Universidad Autónoma de Madrid
Departamento de Física Teórica de la Materia Condensada

Dynamics of virus assembly and budding: a coarse-grained approach

Tesis doctoral presentada por
Teresa Ruiz Herrero

Programa de doctorado de Física de la Materia Condensada

Directores:
Michael F. Hagan
Enrique Velasco Caravaca

Diciembre 2012

Contents

List of Figures

Abstract

Many animal viruses are enveloped, meaning that they acquire a lipid envelope from their host cell. Enveloped viruses can be divided into two classes. Influenza and most retroviruses, including HIV, simultaneously assemble while budding from the membrane. For others (e.g. alphaviruses and hepadnaviruses) a nucleocapsid first assembles in the cytoplasm and then buds from the membrane. While many aspects of budding for some viruses have been investigated via structural studies or biochemical assays, the mechanisms by which these two classes of budding processes occur, and how they depend on parameters such as protein concentration or bilayer rigidity, are poorly understood. By revealing key features of assembly and membrane deformation processes that are not accessible to experiments, simulations can play a role in addressing this limitation.

This thesis consists on the development and analysis of models of simplified representations of capsid proteins and lipid membrane that retain the essential physics while enabling simulations on length and timescales relevant to capsid assembly and budding. Continuum elastic theory is also used for comparison with the simulations results to asses the validity of the models.

In the first part of this work we examine the budding of an already pre-assembled virus through a lipid bilayer using molecular dynamics simulations, free energy calculations, and an elastic theory, with the aim of determining the extent to which equilibrium elasticity theory can describe the factors that control the mechanism and efficiency of budding. The diiferent types of outcomes are also analyzed as a function of the relevant system parameters.

In the second part, we study the assembly of icosahedral capsids on a model membrane domain using elastic theory and molecular dynamics simulations to analyze the geometrical and dynamical properties of the system. With the exception of one recent equilibrium study, there have been no computational investigations of simultaneous assembly and budding. In the simulations presented here, assembly and budding are driven by capsomer-capsomer binding energy and capsomer-membrane adhesion energy. The requirements for capsomer assembly on a membrane compared to bulk assembly, the system features, and the different types of behaviors as a function of the strength of the adhesion energy and the domain line tension and size are investigated. The dynamics of the system are also analyzed by studying how the rate of capsomer arrival to the budding site affects the system behavior.

Resumen

Muchos virus presentan una envoltura lipídica que adquieren de la membrana de la célula infectada durante el proceso de gemación. La forma en la que se adquiere la envoltura sirve para clasificar este tipo de virus en dos grupos diferentes: los que se autoensamblan en la membrana durante la gemación (mayoría de retrovirus, incluyendo el VIH) y los que se autensamblan en el citoplasma antes de adherirse a la membrana.

Aunque muchos aspectos de la gemación han sido estudiados para diferentes tipos de virus a partir de datos estructurales y experimentos bioquímicos, no existe una teoría generalizada del mecanismo por el que estos dos tipos de gemación ocurren y cómo dependen de los diferentes parámetros del sistema, como la concentración de proteínas o la rigidez de la membrana. La simulación de estos procesos puede permitir superar estas limitaciones y describir características importantes sobre el autoensamblaje y la deformación de la membrana durante la gemación que no son accesibles experimentalmente.

Esta tesis se basa en el desarrollo y análisis de modelos simplificados para representar cápsidas víricas y membranas lípidicas que mantengan las propiedad físicas fundamentales al mismo tiempo que permiten la simulación en las escalas de espacio y tiempo características del problema. Los resultados de las simulaciones serán comparados con descripciones obtenidas con teorías elásticas para comprobar la validez de los modelos propuestos.

En la primera parte del trabajo, estudiamos la gemación a través de una bicapa lipídca de una cápsida previamente ensamblada usando simulaciones de dinámica molecular, cálculos de energía libre y teoría elástica para determinar hasta qué punto las teorías elásticas pueden describir los factores que controlan los mecanismos de gemación. Se han analizado también los posbiles comportamientos del sistema en función de los parámetros de relevancia.

En la segunda parte, estudiamos el autoensamblaje de cápsidas icosaédricas en un dominio de membrana a través de teoría elástica y simulaciones de dinámica molecular para analizar las propiedades geométricas y mecánicas del sistema. Para ello, utilizamos un modelo de mínimos donde los enlaces entre capsómeros y la adhesión entre capsómeros y membrana constituyen las interacciones responsables para el autoensamblaje y la gemación. Exceptuando un trabajo reciente que investiga los estados de equilibrio de este tipo de sistem, no ha habido estudios computacionales de procesos en los que el autoensamblaje y la gemación ocurran simultáneamente. En esta parte hemos estudiado los requistos especiales que tiene que cumplir el modelo para que los capsomeros se autoensamblen en una superficie en lugar de hacerlo en volumen. Tambíen hemos analizado las características del sistema

y los diferentes tipos de comportamientos que se pueden observar en función de la energía de adhesión, el tamaño del dominio y la tensión de línea de su contorno. Finalmente, hemos analizado las propiedades dinámicas del proceso estudiando cómo afecta la tasa en la que los capsomeros llegan al dominio al comportamiento del sistema.

Preface

Viruses have historically been regarded mainly as infectious agents because of their parasitism relation with any other form of life. First researches in virology aimed to control viral diseases that seriously affected human health and economy. In human history, major epidemics on people, cattle and crops have repeatedly altered demographic and politic order. Still nowadays, the battle against diseases is the most important driving force in virology research.

This anthropocentric point of view sometimes hides the amazing viral world. Viruses are the most abundant type of biological entity and are found in every ecosystem on Earth. Their wide host range makes them essential regulators in most ecosystems where infection keeps the different populations at equilibrium. Viruses are also considered a natural mean of transferring genes between different species, which increases genetic diversity. Moreover, their great simplicity compared to eukaryota, prokaryota and archea makes them ideal living laboratories for the understanding of cellular biology. Recently, biotechnology and nanoscience have used them as vehicles to introduce genes or drugs into cells.

In this work, we investigate the dynamics of assembly and budding of virus particles. The difficulties of studying this process in vitro, together with the high number of components that take part on it, have so far kept many of its features unknown. In this context, the system modelization can help to shed light on the basic properties of the process. We develop minimum models to focus on the simulation of virus assembly and budding, distinguishing between two different cases: processes where assembly and budding proceed consecutively and those where they occur simultaneously.

The work presented in this thesis has been done with a physics background in a physics department. Despite viruses and membranes have always been a playground for physicists motivated by the interesting properties of these many particle systems and their mathematical modeling, this work does not pretend to consider its subject as a pretext for dealing with a beautiful and astonishing complex system, but as an end in itself. The physical and mathematical concepts are supposed to be the means to understand the process from a different perspective to that adopted by classical biology. This point of view is kept during the whole text and is exemplified in the chapter organization.

The thesis is divided in two main blocks: in the first part (chapters 1 to 3) the properties of the system and the methods used for its analysis are described, and in the second (chapters 4 and 5) the results of this study are presented. Chapter 1 deals with the biological context,

placing emphasis on the main properties of membranes and viruses. The author does not expect to give an exhaustive or very detailed biological description, but tries to emphasize the general properties that play important roles in the modelization of virus assembly and budding.

Chapter 2 describes the computational and theoretical methods required for this study and Chapter 3 summarizes different models that have been proposed and used to study membranes and viruses separately. Special attention is given to the models our work is based on.

In Chapter 4 and 5 the study of assembly and budding is presented. In Chapter 4 we focus on the case where a pre-assembled particle buds out of the membrane and in Chapter 5 the process of simultaneously assembly and budding is analyzed.

Finally, the thesis concludes with a general discussion of the results presented in the previous chapters.

1 Biological framework

1.1 Membranes

Cell membranes constitute one of the fundamental structural and functional components of living organisms. Some theories support the hypothesis that the first cell was originated when self-replicating RNA got surrounded by basic biomembranes, and thus, became separated from the outer environment. Inside the membrane, the self-replicating RNA would have stayed together instead of freely diffused; its higher concentration would have improved the self-reproduction rate and accelerated evolution. Further compartmentalization within the cell would have eventually arisen and not only RNA higher concentrations would have been promoted, but any specific thermodynamical or chemical conditions different from those in the surroundings could have been easily maintained, allowing the diverse cell components to specialize in a wide choice of milieus, independently of the conditions of the external environment.

Apart from acting as barriers, which not only separate the cell from the extra cellular medium, but also compartmentalize its interior, biological membranes have evolved into complex aggregates that play an important role in many physiological functions such as ion and molecular transport, cell recognition, cell adhesion and signaling or regulation of chemical gradients.

Cell membranes are mainly formed by phospholipids and proteins, while proteins are responsible of the regulation of most of their physiological functions, phospholipids confer them flexibility, permeability and fluidity. Those characteristics make them the ideal candidates as semipermeable barriers and structural components, while enhancing the protein activity. These unique properties also represent an interesting area of study both for fundamental and materials science.

1.1.1 Membrane Structure

In 1972 Singer and Nicolson presented the Fluid Mosaic Model to describe the cell membrane as *an oriented, two-dimensional, viscous solution of amphipathic proteins and lipids in instantaneous thermodynamic equilibrium* [1] (Fig. 1.1). This model was able to explain previous experiments and predicted new properties that were not known in those days. Nowadays, this model has been extended to describe a more heterogeneous and compartmentalized membrane

interconnected with other cellular components.

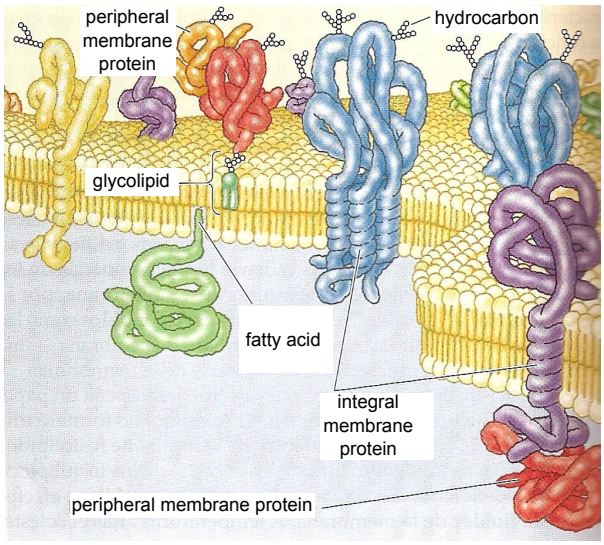

Figure 1.1: Representation of the Fluid Mosaic Model. The membrane matrix is a fluid lipid bilayer formed by phospholipids. Proteins are inserted or attached to one of the layers, and both lipids and proteins are free to diffuse. Figure adapted from [2]

The membrane matrix is a lipid bilayer formed by a double layer of phospholipids oriented so that their polar heads reside on the outer surfaces while the hydrocarbon chains are buried in the inside. The outer part of the bilayer is formed by hydrophilic residues totally hydrated and is typically around 0.8-0.9 nm thick. Next to it there is an intermediate region of $0.3nm$ where the molecule that binds the amphiphile head with the tail lies, this region is only partially hydrated [3]. The hydrophobic core of the bilayer, constituted by the hydrocarbon chains, is typically 3-4 nm thick, although this value varies with chain length and phospholipid chemistry [4, 5] (Fig. 1.2). The water concentration drops from 2M on the head group side to nearly zero on the tail side [6]. Core thickness also varies significantly with temperature, particularly near a phase transition.

One of the most important implications of this structure is the ability of the membrane to easily adapt to different internal and external factors. The fluid state allows its components to freely diffuse, adapting thus to chemical gradients, mechanical stresses or thermodynamical changes. Apart from allowing protein and lipid diffusion, the fluid state also allows the diffusion of some types of small molecules across the bilayer. The two dimensionality represents an advantage for many processes, since it decreases the average time for two membrane components to diffuse and find each other. All this makes the membrane a strong but elastic semi permeable surface and an optimal site for many physiological processes to occur.

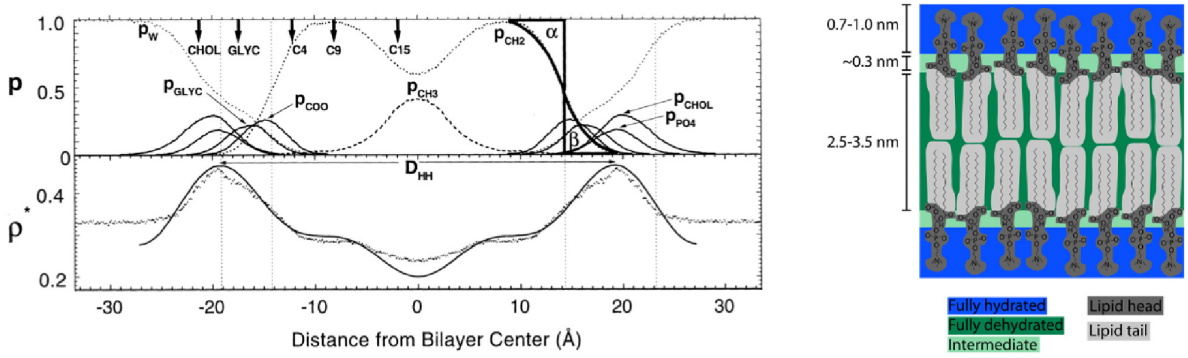

Figure 1.2: Bilayer profile. Left: Representations of structure of Dipalmitoylphosphatidylcholine (DPPC) bilayers in the L_α fluid phase. In the upper figure, the probability distribution functions p for different component groups obtained in simulations is represented. In the lower figure, the electron density profile $\rho*$ from X-ray studies (solid line) and from simulations (dots) are depicted. Figure adapted from [3]. Right: diagram showing the extent of water penetration into a lipid bilayer

1.1.2 Membrane composition

Biological membranes are ubiquitous in every eukaryotic and prokaryotic cell. Depending on the organism, the tissue and the organelles, biological membranes present different specific compositions. However, all of them share two principal ingredients: lipids and proteins.

Lipids

Lipids play important roles in cell physiology as energy stores, signaling molecules and messengers in signal-transduction pathways, and as structural components of membranes, where they not only act as building blocks, but also confer the membrane with unique properties that make it a highly adaptive powerful barrier, as well as an optimal place for many cellular processes to occur.

In their simplest form, lipids constitute fatty acids, long hydrocarbon chains of various lengths and degrees of unsaturation terminated with carboxilic acid groups (COO^-). Unsaturated fatty acids contain at least one double bond within the fatty acid chain, while saturated fatty acids have no double bonds between the carbons atoms of the hydrocarbon chain. Fatty chains are formed by non polar C-H bonds, that do not exhibit interactions with water, so **hydrocarbons are water-insoluble biomolecules**. This hydrophobic property is essential for the formation of membranes.

The length of the chain and the number of unsaturations determine the properties of the fatty acids and of their derivatives. Short chain lengths and unsaturations enhance the fluidity of fatty acids [7].

Phospholipids are the most abundant type of lipids in cell membranes, and are responsible

for most of their singular structural properties. Glycolipids and cholesterol are also lipids that can be found in biomembranes, and enhance their adaptive ability.

Phospholipids form the basic structure of the membrane: the lipid bilayer. They are organic molecules mostly characterized by their **amphipathic nature**, with a hydrophilic polar head and a hydrophobic non-polar tail. They are constituted by four different components: one or more fatty acids, a macromolecule to which they are attached, a phosphate, and a small polar group bonded to the phosphate (Fig. 1.3).

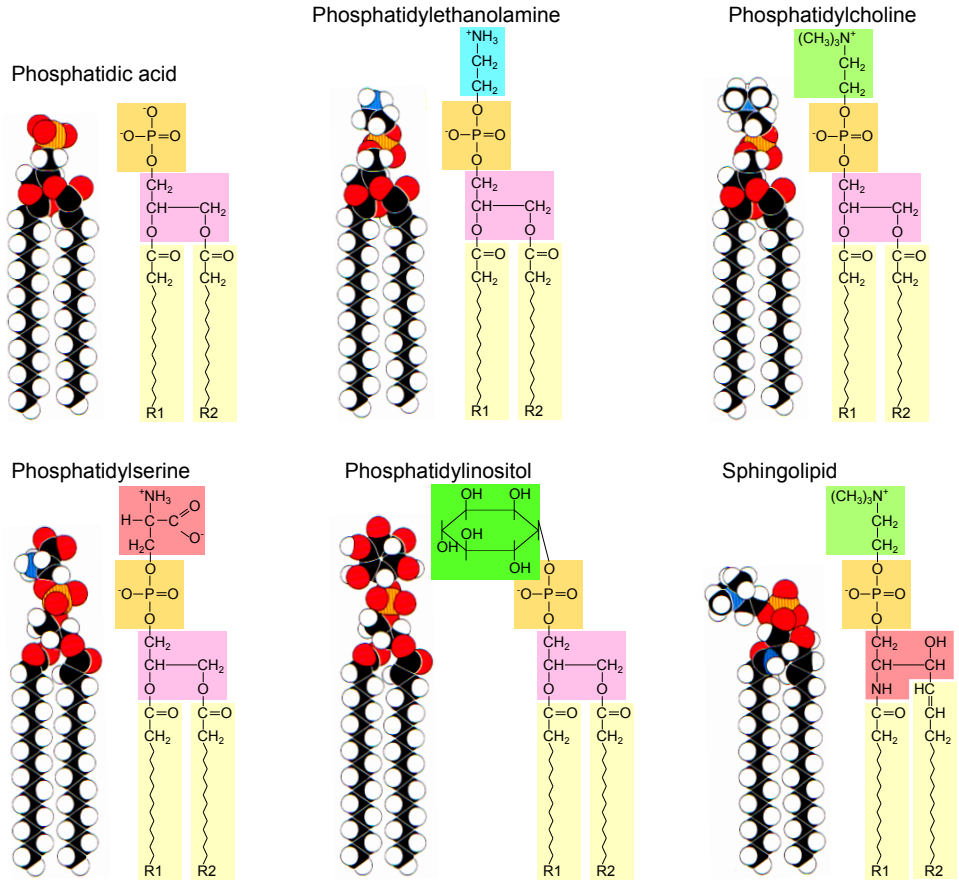

Figure 1.3: Phospholipid structure. Two hydrocarbon chains are linked to a polar head via a glycerol in the case of phosphoglycerids and a serine in sphingolipids. The head group is constituted by a phosphate and a small polar molecule different for each type of phospholipid. Figure adapted from [2]

The hydrocarbon chains show hydrophobic behavior, while the rest of the molecule is hydrophilic. The macromolecule that bonds the tails to the head is often a *glycerol*, and the phospholipids containing this component are called *phosphoglycerids*.

The simplest phosphoglycerid, called *phosphatidate* (PA), has no polar molecule attached to the phosphate group. It is only found in small quantities in membranes, but it is an important precursor for the biosynthesis of many other lipids. Most of the *phosphoglycerids* are

derived from PA by the formation of an ester bond between the phosphate group and the hydroxyl group of an alcohol. The typical alcohols moieties of phosphoglycerids are the amino acid serine, ethanolamine, glycerol and inositol, and the phospholipids they constitute are widely found in biomembranes. The most common are phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylserine (PS) and phosphatidylinositol (PI). Sphingomyelin (SPH) is a phospholipid found in animal cell membranes, that is not derived from glycerol, and instead, it has an amino group.

Since there is a wide range of constituting blocks, as different polar groups or fatty acids with diverse length and variable saturation degree, **phospholipids appear in a wide variety of forms**, with some of them being specific of a tissue or even an organelle (Fig. 1.4), erythrocytes for example have more than 100 different types. Phospholipids found in archeas differ strongly to those in bacteria and eukaryota [8]. In archeas, the hydrocarbon chains are bonded to the glycerol by ether rather than ester linkages, and the fatty chains are branched rather than linear, those differences make the membrane lipids in archeas more resistant to hydrolysis and oxidation, something that can help them to survive in extreme conditions, such as high temperature, low pH or low salt concentration [7].

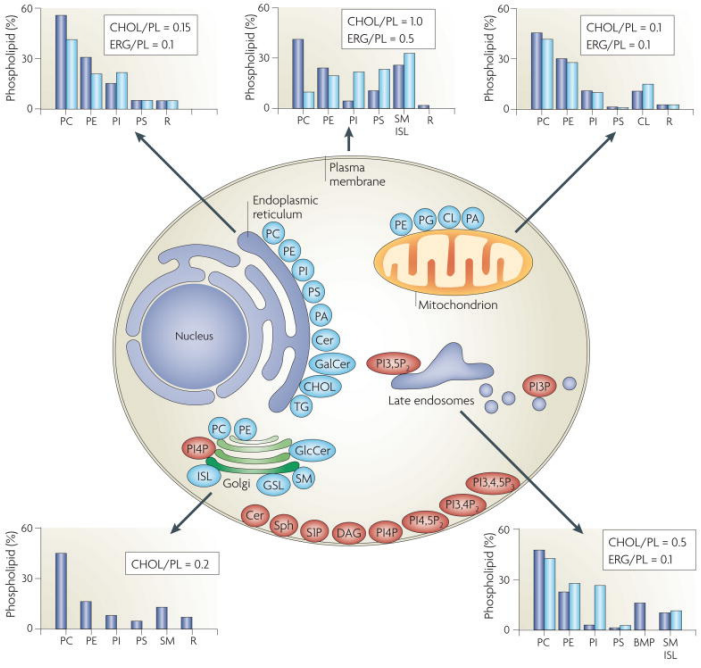

Figure 1.4: Phospholipid diversity in cell membranes. The lipid composition of different membranes varies throughout the cell. The lipid compositional data are expressed as a percentage of the total phospholipid (PL) in mammals (blue) and yeast (light blue). As a measure of sterol content, the molar ratio of cholesterol (CHOL; in mammals) and ergosterol (ERG; in yeast) to phospholipid is also included. Figure adapted from [9]

Glycolipids have a similar structure to phospholipids, with two hydrocarbon chains bonded

to polar groups, formed by carbohydrates. They also present amphipathic behavior. Glycolipids are found in small amounts in the membrane but play important roles as surface membrane receptors and are involved in cell-cell recognition and antigenicity [10] (Fig. 1.5).

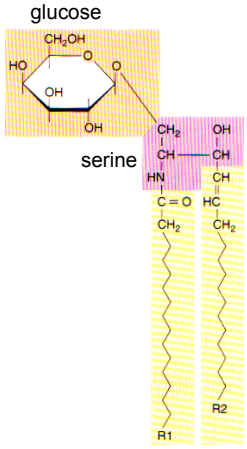

Figure 1.5: Glycolipid structure. Two hydrocarbon chains bond to a head polar group constituted by serine and carbohydrates. Figure adapted from [2]

Cholesterol is only found in eukaryota and is highly presented in animal cells at varying degrees. It constitutes almost 25% of the membrane lipids in certain nerves cells. It is formed by four linked hydrocarbon rings attached to a hydroxyl group (Fig. 1.6). The hydrocarbon rings are highly hydrophobic so they get embedded inside the inner part of the membrane alongside with the hydrocarbon chains, while the hydroxyl group interacts with the phospholipid head groups. Cholesterol is found in all animal cells, and plays an important role modulating membrane fluidity over the range of physiological temperatures by interacting with the phospholipid hydrocarbon chains. Since it is buried inside the bilayer, and due to its non linear structure, it hinders the mobility of the phospholipids tails , reducing the membrane fluidity and its permeability to small molecules. However, at low temperatures, it acts in the opposite way; the interaction with the fatty chains prevents them to tightly pack in an ordered crystal and solidify, so the membrane fluidity is kept at an optimal value even for these regimes [2].

Proteins

While lipids are the main structural components of membranes, proteins are responsible for most membrane processes. Functions of membrane proteins include cell-cell contact, surface recognition, cytoskeleton contact, signaling, enzymatic activity, or transporting substances across the membrane. Protein, as phospholipid, content differs substantially in membranes.

Integral proteins are permanently bound to the membrane, transmembrane proteins span across the membrane, once or more, while other proteins are permanently attached to it from

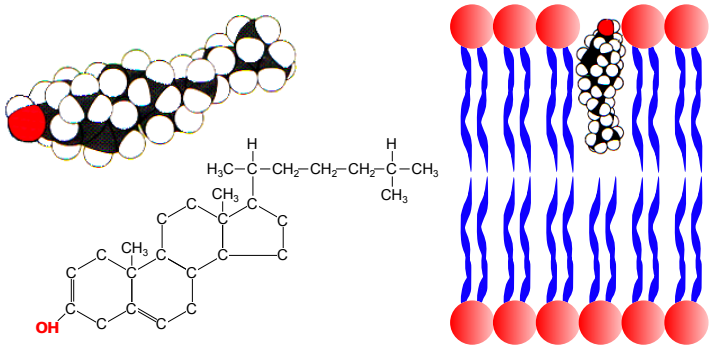

Figure 1.6: Left: Cholesterol structure. A hydrophobic ring with a hydrophilic hydroxyl group. Right: Cholesterol inserts into the membrane with its hydroxyl group oriented toward the aqueous phase and its hydrophobic ring surrounded by fatty acid tails of phospholipids. Figures adapted from [2]

only one side (Fig. 1.7). The transmembrane regions of the proteins are either beta-barrels or alpha-helical, and they are often related to transport mechanisms across the membrane. Porins for example, as their name suggest, form pores or channels specific to different types of molecules across the outer membrane of bacteria, mitochondria and chloroplasts. They are built from β strands, with amino acid composition such that polar and nonpolar residues alternate along the strands. In this way, the outer part of the beta barrel is covered by nonpolar residues that interact with the nonpolar lipid membrane, whereas the polar residues lie in the inner part and interact with the aqueous channel.

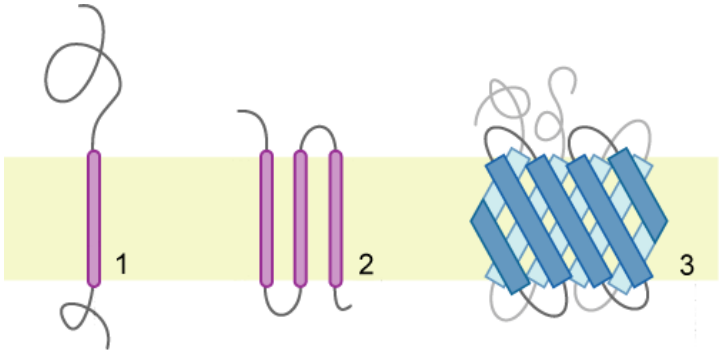

Figure 1.7: Schematic representation of transmembrane proteins: 1. a single transmembrane α-helix, 2. transmembrane α-helical protein crossing the bilayer multiple times, 3. a transmembrane β-sheet protein. The membrane is represented in light brown.

Peripheral proteins however, do not interact with the hydrophobic core of the lipid bilayer. Instead, they interact only temporally with the external part of the cell membrane, either directly by bounding to the head groups of lipids or indirectly by binding to the integral membrane proteins. They can be located in the inner or outer part of the plasma membrane

and are responsible for most of the functions described above (Fig. 1.8).

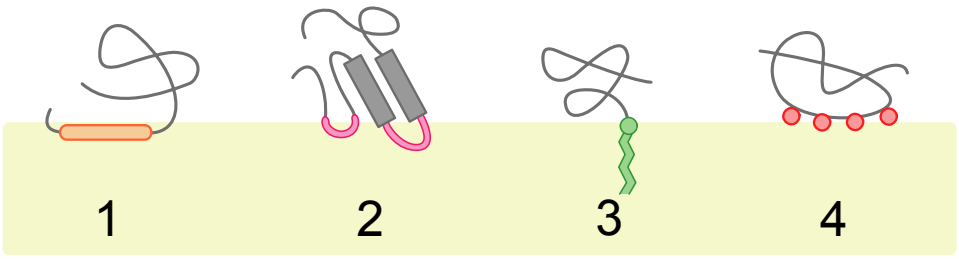

Figure 1.8: Schematic representation of the different peripheral proteins depending on the interaction with the membrane: 1. interaction by an amphipathic α-helix parallel to the membrane plane, 2. interaction by a hydrophobic loop, 3. interaction by a covalently bound membrane lipid, 4. electrostatic or ionic interactions with membrane lipids

1.1.3 Membrane properties

Hydrophobic effect and Self-assembly

Some molecules, in specific conditions, tend to self assemble spontaneously and form an organized structure from an originally disordered system. This is the mechanism responsible for the formation of very different systems such as colloids, liquid crystals, micelles or lipid bilayers. The common characteristic of these systems is the stabilization of the aggregates through non covalent interactions. In the case of membranes, those interactions do not only allow their formation via self assembly of their individual components, but are also responsible of their unique features.

Self-assembly from disordered systems involves an entropy loss that is balanced by a reduction in internal energy due to the interactions that are formed in the new configuration [11]. These final structures are in equilibrium or metastable states, so they lie in a minimum of the system free energy.

Membranes stand apart from other solid components in the cell and from most conventional solid materials due to their fluidity. This is because the forces that hold the amphiphiles together are not of covalent type but of weaker nature. In the presence of water, some molecules tend to aggregate in order to minimize the contact area with the solvent. This is called hydrophobic effect and is the responsible for the formation of membranes [12].

Water is the universal solvent of life so most living organisms have adapted to its special properties. Water molecules are formed by a central oxygen atom linked by covalent bonds to two hydrogen atoms oriented at an angle of 104 ° to each other (Fig. 1.9). Since the oxygen attracts the electrons much more strongly than hydrogens do, the molecule is polarized, and as a result, a net positive charge appears on the hydrogen atoms and a net negative charge on the oxygen. These dipoles induce an electric interaction between the molecules.

However, water dipoles are of a special kind, and water cohesion is sustained by more than normal electrostatic dipole-dipole interactions. In the case of water molecules, each dipole is formed by an hydrogen atom that has only one electron. When this electron is attracted by the oxygen, the hydrogen ends up as an almost naked proton. The electric field at short distances of the naked proton is much more intense than in the case of other atoms that retain the rest of the electrons. Due to these strong interactions, the molecule orients itself in such a way that the hydrogens point to oxygen atoms of neighbor molecules, and a tetrahedral net of water molecules is formed. This net is reproduced throughout the bulk so every water molecule is H-bonded with up to four other molecules.

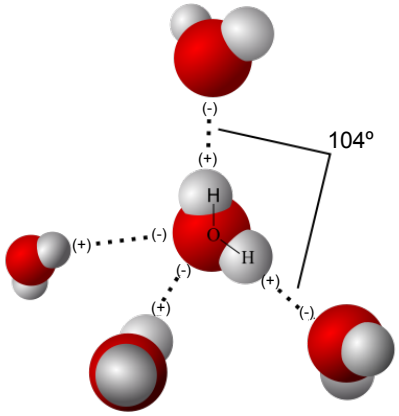

Figure 1.9: Schematic representation of a water molecule forming hydrogen bonds. The angle between the hydrogens of a molecule is 104 °. The bonds between oxygen and hydrogens in such geometry form tetrahedral networks.

The energy of these interactions is about $8k_BT$ stronger than that of polar interactions, but still an order of magnitude smaller than that of covalent bonds. Specifically, the energy change when two molecules of water bond together is about $-9k_BT$, while the generic van der Waals attraction between any two small neutral molecules is about $0.6 - 1.6k_BT$ and chemical bonds are on the order of hundreds of k_BT [13].

However, the specific properties of water, and the hydrophobic effect itself, do not appear from the nature of these bonds alone, but also from the network they form [14]. Even liquid water remains partially ordered, it adopts a compromise between the energetic drive to form a lattice and the entropic drive to disorder. H-bonds are quite transient, with typical timescales only about a picosecond, but on average there is about 3.5 of them per water molecule at room temperature [15], close to the 4 bonds in the ideal tetrahedral network. Liquid water high boiling point is due to the high number of hydrogen bonds each molecule can form relative to its low molecular mass. Due to the difficulty of breaking these bonds, water has a very high boiling point, melting point, and viscosity compared to other similar liquids not formed by hydrogen bonds.

Solubility is highly related to the disruption and rearrangement of this network [16]. Small polar molecules easily dissolve in water, just by forming H-bonds with the solute molecules without disrupting the tetrahedral network. In the case of nonpolar molecules, things are more difficult. They cannot offer any site for H-bonding, so they cannot participate in the network. If the nonpolar molecule is quite small, the hydrogen bonding network distorts to accommodate the nonpolar molecule [17], so that most of the H-bonds are maintained. In this way, there is not an increase of internal energy when dissolving a small nonpolar molecule, but there is a reduction in entropy since the lattice needs to rearrange in a specific way to include the solute molecule. This accounts for an extra cost for dissolving a small nonpolar molecule in water, although the dissolution is still favorable because the van der Waals interactions with water molecules are more favorable than between nonpolar molecules.

In the case of big molecules, the scenario differs importantly. The distortions in the hydrogen bonding network are not big enough to include these molecules, so the only way to fit them implies a disruption of the tetrahedral network and therefore, a decrease in the number of H-bonds. This translates again in a cost in free energy, although in this case, the entropy remains nearly constant and the internal energy increases.

This free energy cost is responsible for the poor solvation of nonpolar molecules in water, a phenomenon usually called hydrophobic effect. Although it has been very commonly referred as an entropic interaction, hydrophobicity cannot be considered a pure entropic effect, since depending on the size of the solute molecule and the temperature, the water network rearrangement can imply an entropy or an internal energy cost [16].

For small molecules, the solvation energy is proportional to their volumes, while for big molecules, it is only proportional to the surface area, since only the very first layer of water molecules is disrupted by the presence of a solute molecule. Assembly of small molecules into an aggregate crosses the threshold between scaling with volume and area, and consequently leads to a lower total free energy. This explains why hydrophobic molecules tend to aggregate in water, and why hydrophobicity is a driving force for self assembly of nonpolar molecules.

Micelles, bilayers and vesicles

When placed in polar solvents, phospholipids present an interesting behavior and a whole field of study. Although these features are common to many molecules, together with their central role in membrane physiology, make them the ideal molecules on studies of self assembly and collective behavior. These interesting features come from their nature: phospholipids are molecules with two different tendencies; their head, a polar residue, is hydrophilic, while the tail, usually a hydrocarbon chain, is hydrophobic. They share this double nature with a wider group of molecules, usually called **amphiphiles**. When those molecules are placed in a mixture of oil and water, they can satisfy both behaviors and reduce the oil-water interface

tension by lying on the contact surface. On the other hand, when placed in water, a large number of possibilities for aggregation arises, with a general structure where the polar parts of the amphiphile face the water while the hydrophobic tails are arrested inside the aggregate core. The different ways to assemble are a result of an optimization of the hydrophobic effect with a variety of intramolecular interactions in combination with geometric constraints. The effects of the electrostatic interaction between the head groups, or the lack of it, the structure of the lipid tails, and the effect of the solvent properties, such as temperature, pH or salt concentration, can be grouped in a phospholipid effective shape. The types of aggregates that are formed by these effective lipids will depend mainly on the relation between the volume v of the hydrocarbon chain, the maximum chain length l_c, and the head group area, a [18] (Fig. 1.10).

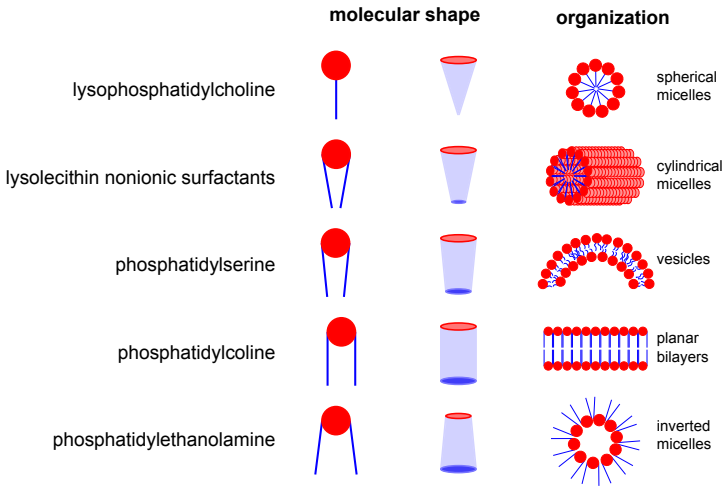

Figure 1.10: Different kinds of phospholipid aggregates depending on the effective molecular shape.

Amphiphiles with big head areas and small chain length and volume have an overall conical shape, and they arrange easily in spherical micelles. The big size of the head area easily forms a shell that can bury inside the small volume of the hydrocarbon chain. This is usually the case for single-chained amphiphiles with charged head groups that interact repulsively with the polar groups of the surrounding molecules, leading to large head-group areas. The hydrocarbon chains are usually short and its length determines the micelle radius.

Amphiphiles with smaller head-groups cannot pack into spherical micelles but form cylindrical structures. This is the case for single-chained molecules with uncharged head groups, or charged head groups in high salt solvents that screen the electrostatic interaction. While spherical micelles have a maximum aggregation number, above which the interior of the micelle cannot be well shielded from the solvent, cylindrical structures are polydisperse in length and their mean aggregation number increases proportionally with the concentration.

In the case of phospholipids, the hydrophobic tail is formed by a double chain, which increases the volume occupied by the hydrocarbons. The overall shape of the phospholipids

is usually cylindrical so they cannot fulfill the geometrical constrains of micelles, and they tend to form bilayers. When bilayers are big enough, or the amphiphile head group areas are slightly bigger than the tails, they can close and form spherical vesicles so the free energetic cost of maintaining the hydrophobic parts in contact with the solvent at the edges of the bilayer is eliminated. This is the case of cellular membranes, vacuoles, lysosomes and other numerous types of vesicles found inside cells.

Membrane Fluidity, Phase Behavior and Rafts

Phospholipids are in a liquid crystal phase, so the heads are not confined to specific positions in the lattice, and the hydrocarbon chains are not rigid and fluctuate freely, this structure is responsible of the fluid behavior shown by membranes. Fluidity is a measure of lipids mobility and, as previously mentioned, is one of the key properties of membranes , since it allows free lateral diffusion of most of their components, which enhances multiple membrane functions. The lateral diffusion rate of lipids in biological membranes is about $1\mu m^2/s$ [7], which means that a molecule can travel from one end of a bacterium to the other in one second, this is 100 times slower than the diffusion of a protein in solution. The fast diffusion allows rapid conformational changes for better adaptation of the membrane to mechanical stresses, such as pore healing, membrane bending and stretching.

Being components of the cell membrane, proteins can also diffuse at different speeds, ranging from those of phospholipids, to slower ones for bigger proteins, or higher protein concentrations [19, 20]. Protein lateral diffusion is related to important cellular processes such as signal transduction, transport or cell communication. While lateral diffusion of lipids can be quite fast, the switching of lipids from one face of the membrane to the other, a process called flip-flop, is usually very slow, with a rate of one flip flop every several hours [21]. Since proteins are bigger, their flip-flop rate is much smaller, and in some cases, their hydrophobic group is so extensive that they cannot even move from one face to the other.

At temperatures below some critical value, phase transitions between liquid crystal and crystalline states are easily observed in biomembranes constituted by only one type of molecules [22, 23]. The thermal motion of the hydrocarbon chains decreases and they straighten, arranging orderly in a close packed structure.

The transition temperature between gel and liquid state depends on the length of the hydrocarbon chains, the number and location of insaturations in the chains, and the amount of cholesterol in the case of animal cells [24]. Lipid mobility is mainly controlled by Van der Waals interactions between the lipid chains. Longer chains present a wider interacting surface, so the Van der Waals strength is higher and their mobility is reduced. That explains why membranes of shorted-chain amphiphiles remain liquid at lower temperatures. Unsaturations also play an important role, the kink they generate in the chain difficults a close-packed arrangement of the phospholipids, and therefore the fluidity is increased. This

effect can be very strong, decreasing the overall chain length by one carbon usually alters the transition temperature of a lipid by ten degrees Celsius or less, but adding a single double bond can decrease the transition temperature by seventy degrees or more.

Bacteria regulate the fluidity of their membrane by varying the number of unsaturations and the length of the hydrocarbon chains in order to keep it fluid at different temperatures. For example, the ratio of saturated to unsaturated fatty acid chains in E. coli membrane decreases from 1.6 to 1.0 as the external temperature decreases from $42\,^{\circ}C$ to $26\,^{\circ}C$ [7]. Animal cells can use cholesterol to adapt to the changes in temperature in a fast and reversible way without important changes in the phospholipid composition that could damage the cell functionality. The presence of cholesterol has a double influence in membrane fluidity. The alcohol group of the cholesterol interacts with the polar heads of phospholipids, so the cholesterol planar rings get inserted in the surface of the membrane and interact strongly with the upper part of the hydrocarbon chains. This interaction decreases the mobility of the phospholipids that surround cholesterol, and a net decrease in fluidity is observed in the membrane. However, at low temperatures, cholesterol deters the phospholipids to arrange in a close-packed configuration, so the transition temperature to the gel phase is decreased [8]. The presence of cholesterol in fluid bilayers is commonly associated to a third phase state, called liquid-ordered phase, which has intermediate properties between those of gel and liquid crystals. In liquid ordered phases phospholipids are extended and ordered as in gels but their mobility is higher, close to that of liquid crystal membranes (Fig. 1.11).

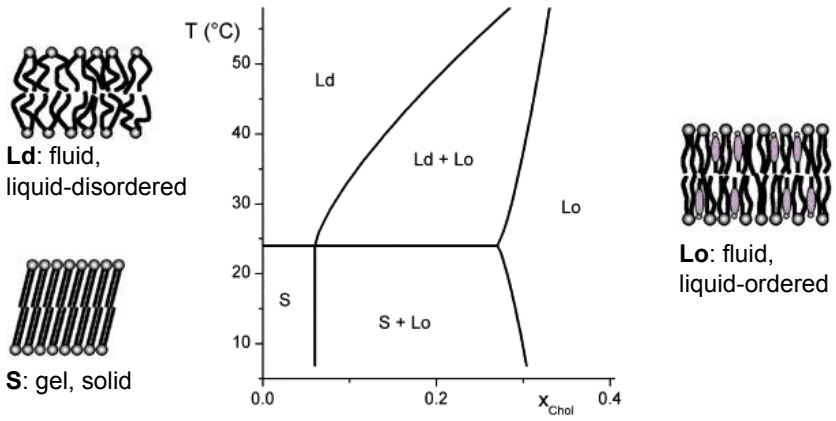

Figure 1.11: Lipid phases. Temperature-composition phase diagram for lipid mixture DMPC/Cholesterol showing the different lipid phases on the system. Lipid phases are denoted as gel phase S, liquid-ordered phase Lo and liquid-disordered phase Ld. Figure adapted from [25, 26]

In mixed lipid bilayers, the transition temperature depends on the nature of each type of phospholipid. When one type of molecule is predominant, the state of the membrane is usually determined by the transition temperature of the most abundant lipid, but in the

case of heterogeneous membranes, with two or more types of phospholipids with similar proportions, phase separation can be promoted, and the gel and liquid crystal states can coexist together. Although the gel state is rare in biological membranes, it has been widely studied in model membranes [27].

Liquid-ordered phase domains are thought to be determinant in many membrane functions. Liquid crystal to liquid-ordered phase separation occurs at $37\,^{\circ}C$ in mixtures of phospholipids, sphingolipids and cholesterol in the same proportions as those of the plasma membrane [28, 29]. These domains enriched in sphingolipids and cholesterol, with an average diameter of 50 nm [28, 30], are called **lipid rafts**. Phospholipids within the rafts are more saturated than in the surrounding membrane, this enhances the close-packed arrangement, but the cholesterol presence maintains the domain fluid in a liquid-ordered phase. The processes which form and regulate lipid rafts are not well understood, and the size range of 10-100 nm is most likely a nonequilibrium characteristic. Local lateral phase separation enables the formation of functional units within the membrane. Lipid rafts have been identified as binding sites for specific types of proteins, sites of assembly and budding of viruses and concentration devices for oligomerization or for proteins acting in signal detection and amplification [30, 31].

Curvature and Tension

As commented in previous sections, the effective lipid shape is not necessarily a straight cylinder, and thus, the default membrane shape is not usually flat. In most cases it presents a spontaneous curvature. Lipids with an overall inverted conical shape form bilayers with positive curvature, while those with conical shapes induce negative curvature. However, phospholipid geometry cannot account either for the strong curvatures observed in certain membrane domains, like those in the Golgi apparatus, the endoplasmic reticulum, and in ruffles and invaginations on the plasma membrane; or for several membrane intermediate shapes observed in cellular processes such as endocythosis, budding or fission. Therefore, those conformations do not appear spontaneously but imply an energetic cost. Membranes resist bending because changing local curvature alters both the head group spacing and the entropy and distribution of the tails to values different from those in equilibrium.

While bilayers containing a single type of phospholipid can only have a small spontaneous curvature, asymmetric bilayers with different population in each of the layers present different membrane curvatures that can induce bending. The transfer of lipids from one leaflet to the other is usually promoted by the transmembrane enzymes, flippases, and maintained by the low rate of phospholipid flip-flop. Membranes, in fact, are structurally and functionally asymmetric [32]. The chemical composition of the leaflets of a bilayer is very different from each other. Sphingomielins and glycolipids for example reside almost exclusively in the outer part of the membrane, while PS is mostly found in the inner part, and cholesterol

appears in both bilayers, although it seems to be more abundant in the outer leaftlet . Protein and lipid assymetry are related to the modulation of cell signaling. Membrane asymmetry produces differences between the membrane bending and stretching moduli of each layer, this condition is exploited by many cellular processes, and plays an important role in the function and regulation of transmembrane proteins [33].

Also proteins can generate curvature when binding or inserting into the membrane [34], specially when transmembrane proteins with conical shapes or peripheral scaffolding force the membrane to adapt to their shape (Fig. 1.12). Cytoskeleton activity contributes as well to membrane shape changes, however, its ability to induce curvature is affected by **membrane tension**.

Membranes not ontly experience bending but also tension when the average area per lipid increases or decreases. Membrane tension controls cell shape and motility [35] and is regulated by the polymerization of the cytoskeleton, the internal pressure due to osmotic flows, or the action of membrane channels. Bilayers resist stretching because increasing the average distance between head groups increases the exposure of the hydrophobic core to water. Strong stretches usually induce pore formation. Bilayers also resist compressions to avoid short range repulsive interactions between the lipids. When the membrane is compressed, it usually buckles out of plane and keeps the total area constant. The magnitude of the resistance to forces applied in different directions is characterized by two elastic constants, the **bending and stretching modulus**.

Finally, heterogeneous lipid populations are responsible of domain formations, that can exhibit local tension and curvature. At the domain boundary, the interaction between molecules of different species produces **line tensions**. Those tensions can control membrane deformation, budding and fission [36, 37]

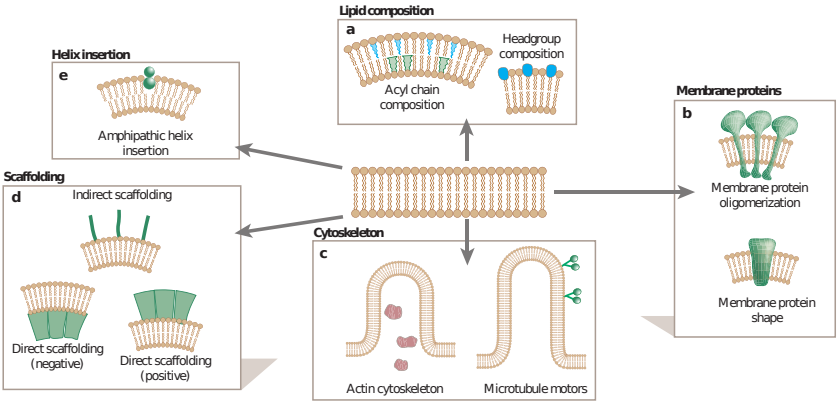

Figure 1.12: Different processes that produce curvature on the membrane. Figure adapted from [34]

1.1.4 Membrane transport

Cell membranes separate the cell from its surroundings and compartmentalize the different organelles, but cells and organelles are not isolated and need to interchange molecules with the outer environment. Membrane specific permeability ensures right molecular concentrations inside the cell. Although the membrane is in a fluid state, most of the biological molecules and macromolecules cannot diffuse through it, so a big family of proteins has specialized in transporting them in and out the cell.

Transport of Small Molecules

Small hydrophobic molecules can diffuse through the cell membrane because they dissolve in the lipid bilayer. The membrane is totally permeable to those molecules, only the difference in the concentration at both sides regulate their free diffusion. Polar or charged molecules however, cannot dissolve in the interior of the lipid bilayer, so they need to cross the membrane through specific paths that do not expose them to the lipid core. Transport proteins form those paths to facilitate diffusion of polar species in the direction of the concentration gradient. These proteins span the membrane and can form open channels specific to each type of molecule, as in the case of sodium channels, or they can just work as gates, allowing the pass of the tagged molecules through conformation changes.

However, not all the molecules need to be transported in the direction of the concentration gradient, in fact, different ion concentrations between the cytoplasm and the extracellular space are usually maintained for the right function of the cell. In those cases, the molecules often cross the bilayer against the concentration gradient. These processes need an external energy source that can be taken from chemical reactions, such as ATP hydrolysis or even from light, as proteins involved in photosynthesis do. In other cases, the transport of free diffusing species is coupled to that of a second type of molecule going against the gradient, the sodium-calcium exchanger is an example of this kind of mechanism.

Transport of Macromolecules: Endocytosis

Crossing the membrane, directly or through protein channels, is not possible for macromolecules or other biological complexes because they are simply too big compared to the phospholipids or the transport protein sizes.

For those particles, endocytosis is the common way to enter the cell. A portion of the membrane invaginates as the particle is progressively engulfed; at the point where it is mostly wrapped, a thin neck is formed that finally pinches off releasing the vesicle in the cell interior. This mechanism is used for nutrient uptake, cell signaling or to remove pathogens and cell debris by the immune system. It is also known that viruses and bacteria often hijack the endocytic pathways to enter the cell [38, 39]. The internalized particles or vesicles range

in size from 50 nm to 5μm.

Several endocytic pathways have been identified for different cell types and cargo species (Fig. 1.13). The first and most studied route for endocytosis is clathrin mediated [40]. Clathrin is a large protein that polymerizes into curved lattices on the inner surface of the plasma membrane, forming coated pits. The transported cargo enters those pits by attaching to specific receptors and is released into the cytoplasm inside coated vesicles when the pits bud into the cell. The scission of the pit prior to budding is usually assisted by cellular machinery.

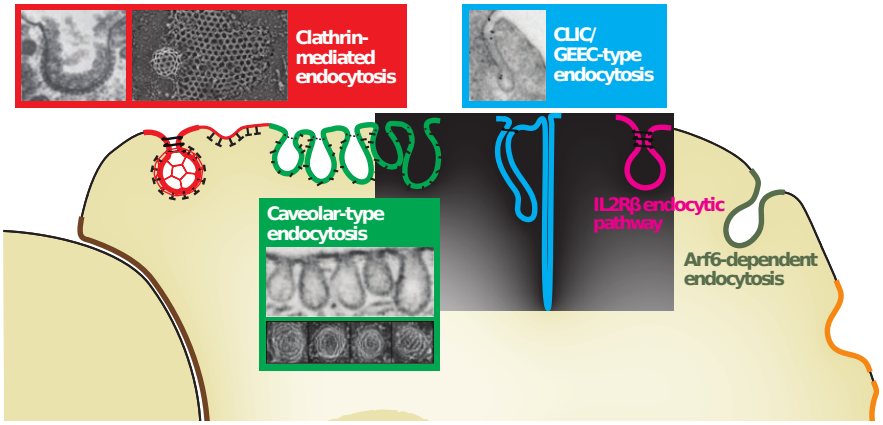

Figure 1.13: Different endocytosis pathways depending on the nature of the uptake and on the membrane site properties. Figure adapted from [39]

Although clathrin mediated endocytosis is used by numerous cells for many different cargoes, other important non-clathrin routes have also been identified [41]. In many cases, endocytosis occurs in membrane sites that present natural highly ruffled regions or invaginations, as caveolae, which are small flask shaped invaginations of the plasma membrane rich in proteins, cholesterol and sphingolipids [39]. Lipid rafts have also been claimed as preferred sites for endocytosis, because of their special qualities: they show spontaneous curvature, protein and lipid specificity and inherent line tensions around them that can enhance budding [30, 42].

Several endocytic pathways have been identified related to specific cell types and to the nature of the intake. Despite of the diversity of endocytic pathways, all of them share a general mechanism that requires the deformation of the plasma membrane to produce highly curved endocytic pits and the scission of these membrane pockets. These processes are usually induced or promoted by proteins and cell machinery.

1.2 Viruses

Viruses are biological entities that need to parasite other organisms to replicate. In their simplest form, they consist of a nucleic acid protected by a protein shell, called capsid (Fig.1.14).

They are usually not classified as a form of life because they lack the machinery needed for reproduction on their own. While it is true that viruses do not have their own metabolism and need a host cell to reproduce, some types of bacteria such as chlamydia and rickettsia [43] are considered living organisms despite of the same limitations. Leaving aside general and non rigorous definitions, viruses are in any case organic structures able to autoreplicate inside the right host and are subjected to evolutionary pressure.

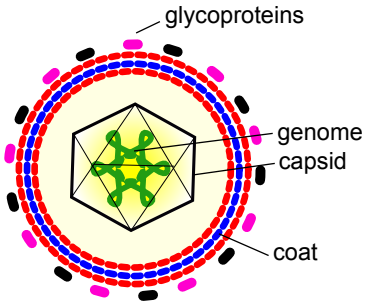

Figure 1.14: Virus structure: a nucleic acid is protected by a protein shell, that in some cases is surrounded by an envelope formed by viral glycoproteins and a portion of membrane from the host cell.

More than 3000 different species of viruses have been identified to date [44] (Fig. 1.15), although it is estimated that there are millions of them in nature [45]. Viruses are found in almost every ecosystem on Earth and they are the most abundant biological entities on the planet, they can infect any type of organism, from animals and plants to bacteria and archea. The high rate of gene modification in viruses via mutation and recombination, can account for their huge diversity. Recombination occurs when the genome of two related DNA or RNA viruses find themselves in the same host and combine together, giving place to a different entity. Mutation happens very often in RNA viruses because the polymerases involved in the RNA synthesis cannot correct errors, a process called proofreading, so mutations accumulate at a rate of approximately $3 \cdot 10^{-4}$/nucleotide per cycle of reproduction. This explains why some viruses get resistant to vaccines in very few generations [46]. On the contrary, DNA polymerases, which are used by most living organisms to replicate their DNA, have the ability to proofread.

Given their great diversity, it is most likely that viruses have not evolved from a single common ancestor, but they have converged from different branches. Since they do not have fossil record and have a high rate of mutation and recombination, it is hard or even impossible to trace back their origin. However, specific characteristics of different viruses suggest more than one possible scenario for their origin. They could have easily evolved from lost of unneeded cell machinery by parasitic cells, from bits of RNA or DNA that escaped from genes of larger organisms and instead of being degraded inside other hosts were able to replicate; or from primitive proteins and plasmid acids at the same time as the first cells

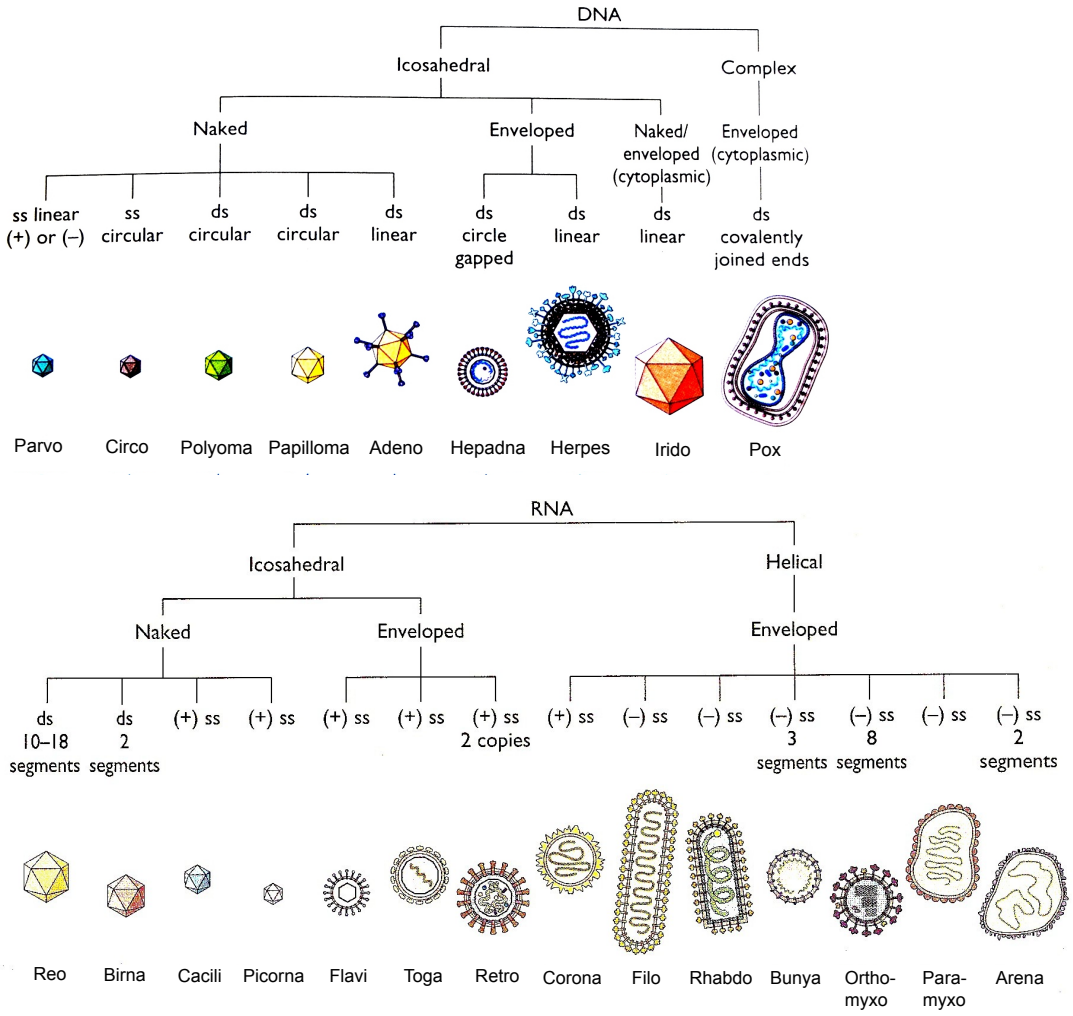

Figure 1.15: Virus phylogeny of 23 representative families that infect vertebrates. A high virus diversity can be observed within these families. Figure adapted from [47]

appeared on Earth. Although none of those mechanisms have been proved, it is very likely that viruses evolved independently from several of those routes [48, 49] .

Despite of their lack of energy-generating and biosynthetic systems necessary for independent existence, they are still more complex than other biological active agents, such as viroids and prions [50, 51]. Nevertheless, their basic structure and the reduced number of processes in their life cycle make them ideal models for genetic and biochemical studies.

1.2.1 Structure

Viruses exist in different sizes and shapes, however their life cycle has common features and thus, all of them are built according to the same general principles. The structure of a virus particle, or virion, consists of nucleic acid surrounded by a protective coat of protein called capsid.

Nucleic acids encode both for viral structural proteins and at least one or more of the molecules associated with biosynthesis. RNA for transmitting information. Although the standard information storage medium in living systems is DNA, **virus genomes** can be encoded either in DNA or RNA molecules. Depending on the type of virus, nucleic acids can be double- or single-stranded, segmented or not, and DNA can also be present in linear or circular configuration. Some viruses also carry together along with the nucleic acids proteins needed for the replication. In the case of RNA viruses, their host cells do not have the enzymes required for RNA synthesis, thus, they carry very often their own enzymes inside their capsid.

The largest DNA genome sizes are bigger than the largest sizes of their RNA counterpart. They range from $3 \cdot 10^3$ base-pairs in hepadnoviruses to $8 \cdot 10^5$ by mimivirus. The biggest RNA viruses however are those from coronavirus family with $3 \cdot 10^4$ nucleotides in length [47]. The lack of proofreading in RNA polymerases can be a reason why RNA genomes cannot exceed a particular size. A mutation rate of 10^{-4} nucleotides per cycle of reproduction means that in a viral genome of 10^4 nucleotides, there would be in average one mutation in every replication. For RNAs longer than $3 \cdot 10^4$, it is very likely that many lethal mutations can occur in a replication cycle. This can be a strong limitation to RNA sizes.

Another limitation to the genome sizes comes from the possibility to make a big enough capsid to enclose the nucleic acid. Viruses present a wide diversity of sizes and shapes, their diameter usually ranges between 20 and 300 nm. **Capsids** are symmetric structures that protect the nucleic acid from the surroundings and act as vehicles transporting the virus genomes from cell to cell, interacting with cell surface receptors and allowing the entry and exit processes. Nucleocapsids are formed only by proteins and nucleic acids, and all the information needed to replicate is encoded in the genetic material.

Genome coding triples weight 10 times more than the proteins they specify, since virus particles contain between 50 and 90% of proteins in weight, there is simply not enough genetic information to build those proteins. What makes viral architecture appealing is the elegant solution nature has found for the capsid design coping with size, shape and genetic information restrictions. Crick and Watson [52] proposed that the presence of identical protein subunits packed in a regular manner would be the only way for the reduced genetic material to encode all the proteins in the capsid. In addition, they claimed that for those asymmetric proteins to build a symmetric capsid, all the subunits needed to be equally bonded so all them interact with their neighbors in an identical, or equivalent manner. In this way the subunits, called capsomers, self-assemble by increasing the number of inter-subunit bonds so that the final structure is in the lowest energy state.

The most common symmetric structures observed in virus particles have cubical or helical symmetry. Filamentous viruses are composed of a single type of capsomer stacked around a central axis in a helical structure (Fig. 1.16). This structure allows the helical nucleic acid to be equivalently bonded to the capsomers. While protein-protein interaction is the driving

force for capsid formation, the interaction between the nucleic acid and the protein stabilizes the structure. The first virus to be identified, the tobacco mosaic virus, presents this type of symmetry with a diameter of 18 nm and a length of 300 nm [53]. The Sendai virus has the same structure and nearly the same diameter, however, it has a length of 1μm [54]. A characteristic feature of the helical structure is that any volume can be enclosed simply by varying the length of the helix, therefore, the final length will be determined by the size of the encapsulated genome.

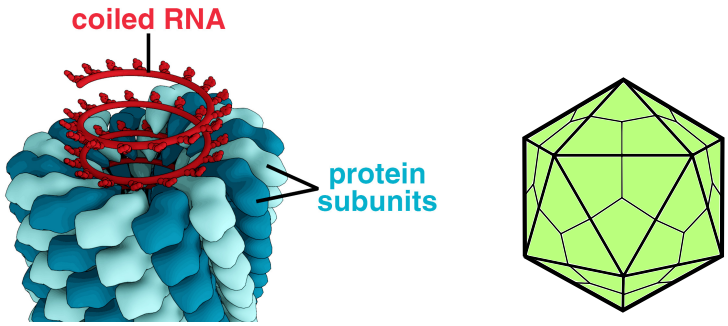

Figure 1.16: Left: helical virus structure. The helical nucleic acid is equivalenty bonded to the capsomers that are also arranged in an helical structure. Figure from Thomas Splettstoesser (www.scistyle.com). Right: Simplest icosahedral virus formed by 60 idential subunits.

Many viruses are spherical instead of elongated. When closely inspected, those spherical shapes turn to be polyhedron with icosahedral symmetry. The simplest icosahedral capsids are built up by using three identical subunits to form each triangular face, thereby requiring 60 identical subunits to form a capsid (Fig. 1.16). Given the limited size of subunits, this geometry is only used in very small particles. Although viruses are usually bigger, satellite viruses that depend on co-infection with fully mature viruses carrying enzymes and proteins they lack, are often constructed in this way. For this geometry, each subunit is related identically or equivalently with its neighbors. The advantage of the icosahedron over other polyhedra with cubic symmetry is that it allows the use of the greatest possible number of identical asymmetric units to build a spherical framework in which they are identically packed. However, it is known that capsids can contain more than 60 capsomers. Caspar and Klug showed that more than 60 subunits can fit in a icosahedron when each subunit occupies a quasi-equivalent position, that is, the bonds between subunits in different regions are similar but not identical [55]. If we try to triangulate a sphere with more than 60 subunits, we will see that the next polyhedra that can be formed is a truncated icosahedron . At some of the vertices of this new icosahedron there are rings of five capsomers, called pentamers, while at other vertices the rings have 6 capsomers, called hexamers (Fig. 1.17). As some of the subunits are arranged as pentamers and the others as hexamers, it is evident that they cannot be equivalently related, although they can interact with each other in a very similar

or quasi-equivalent way.

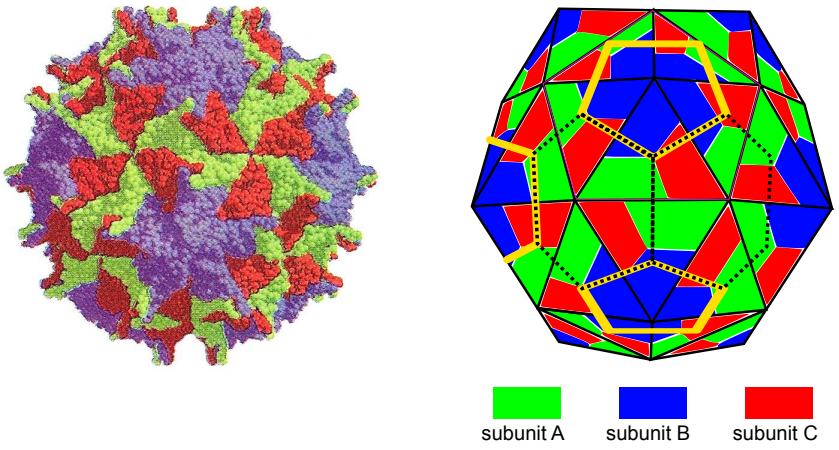

Figure 1.17: T3 capsid. The structure of the Nodamura virus particle determined by X-ray crystalography is shown on the left and it is summarized schematically on the right. The coat protein subunits that occupy A, B and C structural environments are colored blue,red and green respectively. Some of the vertices of the capsid present five-fold symmetry and others six-fold. The capsomer division into pentamers and hexamers is shown in the figure. Figure adapted from [47]

The number of ways in which each triangular face of the icosahedron can be subdivided into smaller, identical, equilateral triangles can be calculated by using the triangulation number (Eq. 1.1), that describes the relation between the number of pentagons and hexagons:

$$T = (h + k)^2 - hk \tag{1.1}$$

where $h > k$ and h and k are the distances between the successive pentagons on the virus surface for each axis (Fig. 1.18). The larger the T-number the more hexagons are present relative to the pentagons. For any T-number, the number of pentagons has to be 12 to obtain a closed structure. T=1 defines a dodecahedron with 12 pentagons, while T=3 would represent a truncated icosahedron. The total number of subunits will be 12 pentagons and $10(T-1)$ hexagons. T number roughly scales with the area of a capsid. As stated above, only satellite viruses and very few small viruses, like the canine parvovirus, show T=1 symmetry. In the case of the satellite viruses, they depend on coinfection by a complementary larger viruses for infection, because they are so small that they cannot accommodate the total genome needed for replication. There are more representatives of viruses with T=3, specially plant viruses, although they are still quite small compared to most animal viruses. Bigger viruses present higher symmetry numbers, adenoviruses for example present T=25 symmetry [56].

Finally, there are some viruses that do not show a pure helical or icosahedral symmetry, and are said to have complex capsids. In many cases, these capsids are a combination of helical and spherical structures, this happens in many bacteriophages that consist of an

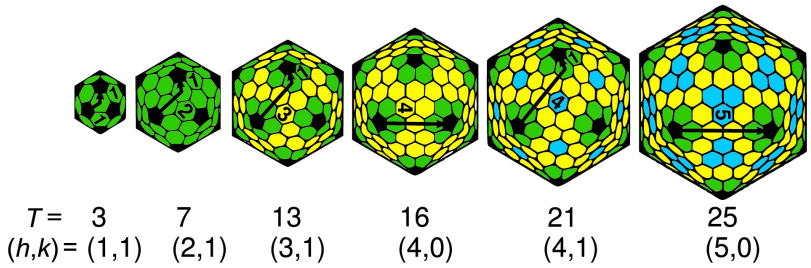

Figure 1.18: T-numbers and their related capsid structures. h and k are the distances between the successive pentagons on the icosahedron surface for each axis. Figure adapted from [57]

icosahedral head bound to a helical tail, which may have a hexagonal base plate with protruding protein tail fibers. In other cases, capsids present long projections, called fibers, at each of the icosahedral vertices, or other different non-symmetric structures.

In many viruses, capsids are surrounded by an **envelope** formed by viral glycoproteins and a portion of membrane derived from the host cell. Viruses acquire the envelope from the plasma membrane while exiting from the host cell or from other internal membranes. The specific lipid composition is variable depending on the nature of the cell membrane, while viral glycoproteins are coded in the viral genome. Viral glycoproteins are integral membrane proteins that identify and bind to receptor sites on the host membrane and mediate fusion, facilitating in this way the entrance of the viral genome inside the host cell. Viral proteins can form spikes in the exterior of the viral particle, giving the enveloped viruses their characteristic shape. Protein cytoplasmic domains make often contact with the virion and play an important role in virus assembly (Fig. 1.19).

1.2.2 Replication Cycle

As parasites, viruses use the machinery and metabolism of the host cell to reproduce, their life cycle consists of all the processes they undergo to enter the host cell, replicate and exit. Due to the high diversity of viruses, those processes vary from one type to another, however, the main features of the replication cycle are common to all of them (Fig. 1.20)

The fist step for most viruses to enter the host cell is the **attachment to the cell surface**, binding between viral capsids and specific receptors allows the virus to get in contact with the cell and usurp the normal cellular transport processes to promote the **insertion of the genetic material** inside the cell. For some enveloped viruses, such as HIV-1, interaction between viral proteins and cellular receptors promotes the fusion of the lipid bilayer of the virus with that of the membrane, resulting in the virus particle being internalized by the host cell. Other enveloped viruses and many non-enveloped viruses enter the cell by receptor-mediated endocytosis. Bacteria and plants have strong cell walls that do not support

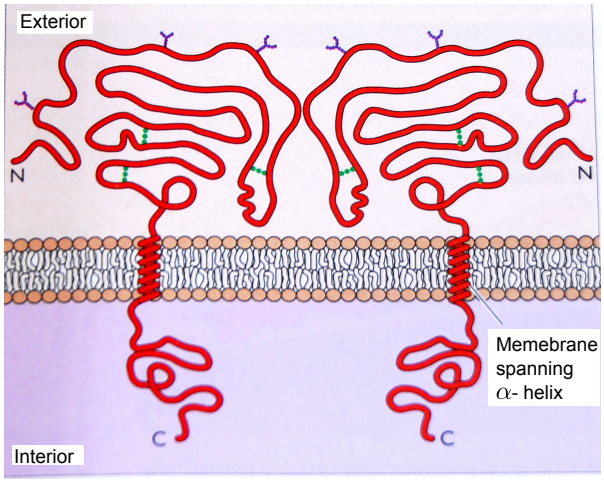

Figure 1.19: The viral glycoprotein has three differenciated parts: an external domain, a transmembrane segment and an endodomain. The external part interacts with the host receptors during entry and the internal part promotes assembly and binding to the membrane before exit. Figure adapted from [47]

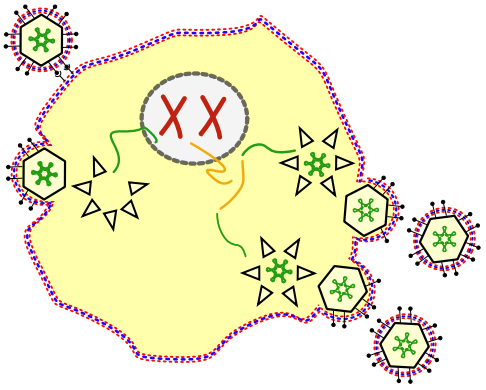

Figure 1.20: Cartoon showing the general steps of virus cycle. 1. Attatchment to the host cell and entry, 2. Replication of proteins and genome, 3. Virus assembly, and 4. exit the host cell by budding

the regular transport processes, instead, the virus must break them to enter the cell. Since bacterial walls are much less thick than plant cell walls, some bacteriophages have evolved mechanisms that inject their genome into the bacteria across the cell wall.

Inside the host cell, once the nucleic acid has been delivered to the right place, it does not need to be protected anymore. During the **uncoating** process the genome is liberated due to the degeneration of the capsid promoted by viral enzymes or host enzymes or by simple dissociation of the protein subunits. This process is poorly understood and appears to occur by a variety of very different mechanisms for different viruses.

Nucleic acids carry all the information needed for the **production of new virus parti-**

cles. Replication involves multiplication of the genome and synthesis of viral proteins. The methods by which the genetic material is replicated and the viral messenger RNA (mRNA) is synthesized for the production of viral proteins vary considerable between the different types of viruses. In DNA viruses, genome replication and mRNA synthesis take place in the cell nucleus and are usually entirely dependent on the host cell DNA and RNA synthesis machinery. In the case of RNA viruses however, replication usually takes place in the cytoplasm and all RNA viruses use their own RNA replicase enzymes to create copies of their genome.

After the virus genome, capsid proteins and other molecules that may be part of viruses are made, they are transported to the assembly sites, where they combine to form virus particles in a process of **self-assembly**. Some capsids may also be assembled from scaffolding proteins which are then modified or even discarded in the final virus particle. The viral genome can have a decisive role in the capsid assembly acting as initiating factor and promoting the capsid formation around itself. In other cases, the capsid forms independently into an inmature particle, and the nucleic acid is introduced in the final steps of assembly.

In the last step of the replication cycle, viruses are **released outside the cell**. In some cases this process is mediated by cell lysis, when the membrane or cell wall undergoes one or several ruptures that kill the cell. Enveloped viruses however, usually exit by budding, during this process the virus acquires its envelope (Fig. 1.21). This mechanism favours virus replication because it mantains the integrity of the host cell membrane, so the cell can continue virus production.

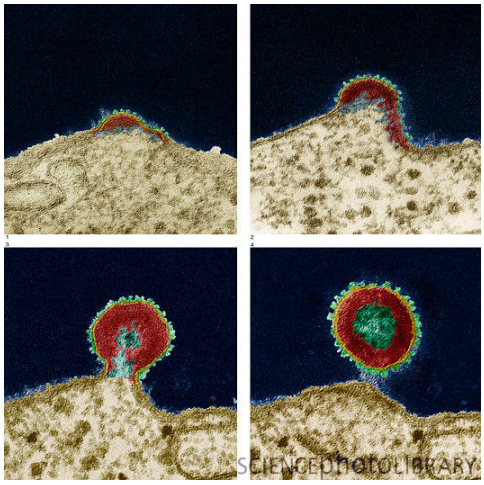

Figure 1.21: Colored transmission electron micrograph of an HIV virus (red/green) budding from the surface of a T-lymphocyte white blood cell. Magnification: x100,000 at 6x6cm size. From EYE OF SCIENCE/SCIENCE PHOTO LIBRARY

After assembly, many viruses are not infectious particles and are often called immature virions. **Maturation** reactions are carried out by virus enzymes and can happen before or

after release, the virus undergoes several irreversible chemical and conformation changes that transform it to an infectious particle able to entry in a new host cell and initiate the reproduction cycle again (Fig. 1.22).

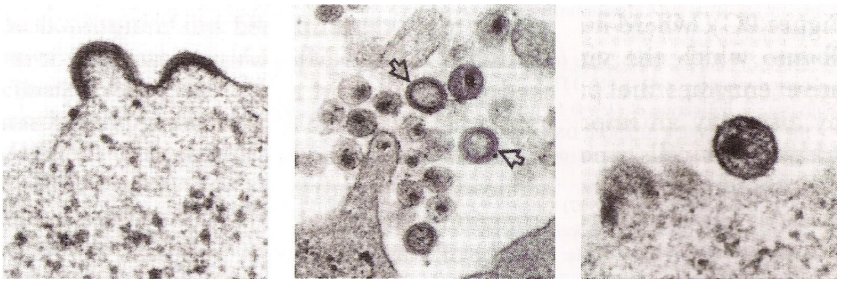

Figure 1.22: Virus maturation. Morphologic changes seen in HIV particles after budding. Left: budding of HIV capsids. Middle: Mixture of both immature (arrowed) and mature HIV particles. Right: Mature HIV virions. Figure adapted from [56]

1.3 Virus Budding

Most enveloped viruses acquire their envelope from the plasma membrane of the host cell when budding out of it as the final step of the replication process. Viruses hijack the common cellular pathways for receptor mediated exocytosis, thus, budding has many features in common with general types of particle endo- and exocytosis. During budding, the viral capsid attaches to the plasma membrane via protein mediated interactions. As the interactions between capsid and membrane proteins increase, the particle is progressively wrapped, until only a thin neck keeps the virus attached to the cell. When the lipid bilayer pinches off, the membrane forming the neck fuses and the enveloped particle is released outside the cell (Fig. 1.23).

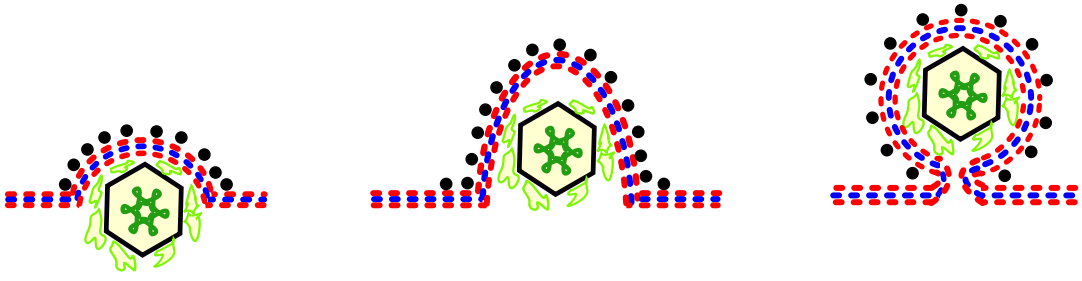

Figure 1.23: Schematic representation of virus budding. First, the particle attaches to the membrane, the virus is progressively wrapped until a thin neck is formed. Eventually the membrane forming the neck fuses, releasing the enveloped virus.

In general there are two types of proteins involved in the attachment between the mem-

brane and the nuclecapsid, although their properties can vary from one virus family to other. In most cases the interaction between capsid and membrane is mediated by transmembrane glycoproteins that have been previously transported to the budding site through the cellular secretory pathway. These proteins can bind directly to the capsid subunits or through a matrix protein that form a layer in the inner suface of the membrane. In some cases budding can also be driven exclusively by matrix proteins that bind to the envelope proteins and to the lipid bilayer. In many retroviruses, the capsid , nucleocapsid and matrix proteins belong to the same polyprotein, which is only cleaved during maturation.

Budding does not always occur in the plasma membrane, some viruses, such as coronaviruses, bud into the endoplasmic reticulum, acquiring cytoplasmic membranes as their envelopes. These virions undergo multiple budding and fusion reactions as they are transported to the plasma membrane: they bud off the endoplasmic reticulum inside a vesicle, fuse into the Golgi complex membrane, and bud off again into a new vesicle that finally fuses into the plasma membrane and releases the envelope particle outside the cell (Fig. 1.24). After the long trafficking, the virus particle retains its original envelope throughout the process [56]. Other viruses follow a similar pathway but from different origins, in the case of the herpesvirus, they assemble in the nucleus and undergo primary envelopment at the inner nuclear membrane [58].

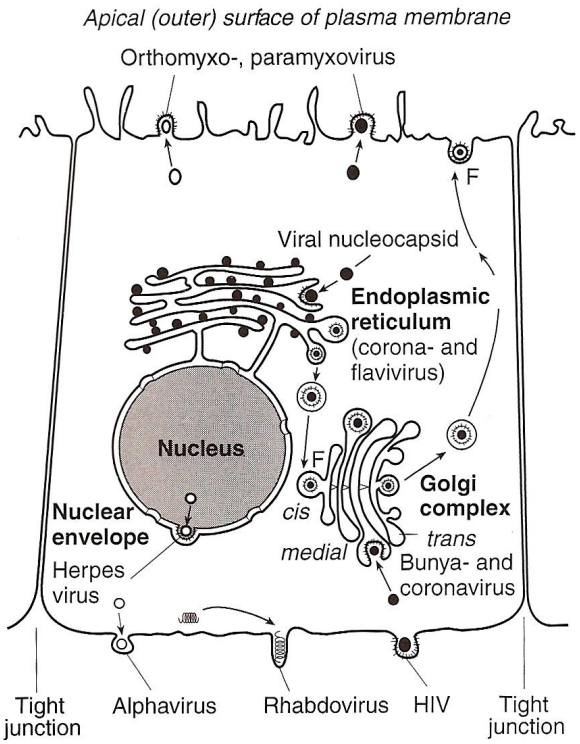

Figure 1.24: Budding of viruses from different families. In some cases multiple budding in internal membranes occurs before the virus exits the cell. Figure adapted from [56]

Budding sites are determined by the destination of viral proteins that enter the cellular secretory pathway, and bind to specific molecules of the host cell. Budding mechanisms can be separated into two general classes, distinguished by whether the acquisition of the envelope follows assembly of the viral particle , or whether both process, assembly and budding, occur simultaneously.

1.3.1 Sequential assembly and budding

In many cases, the assembly of the nucleocapsid and the attachment and budding are spatially and temporally separated. Capsid assembly can occur in the in the cytoplasm, as for the poxvirus [59], or in the nucleus, like in the case of most DNA virus, or some unusual RNA viruses where RNA synthesis occurs in the cell nucleus. The synthesized proteins that are not enclosed by the capsid, are transported to the plasma membrane or any other cellular membrane where budding occurs. The fully assembled nucleocapsid attaches to those proteins that mediate the interaction with the membrane. The membrane progressively wraps the particle as more viral proteins bind to the nucleocapisd until a thin neck is formed that eventually fuses releasing the enveloped virus outside the cell.

Viral membrane proteins not only mediate the interaction between the membrane and virus particles, but also play a crucial role in the maturation of many enveloped viruses. In some cases, interaction between the matrix and capsid proteins is necessary for the formation of virions with normal size, morphology, and with the appropriate concentration of RNA [60]. Other small proteins or peptide complexes not directly related to the binding between the membrane and the nucleocapsid, can be also determinant in the budding mechanisms. This is the case of some viral proteins that hijack the ESCRT (endosomal sorting complex required for transport) protein complex of the cell that enables the invagination and pinching off process during endocytosis [61]. Although some viruses can bud without using the ESCRT complex, induced mutations of the viral proteins that associate with this complex often result in virus particles that cannot be released and remained connected to the cell surface by a thin tether at the final step of budding [62].

1.3.2 Simultaneous assembly and budding

For many retroviruses, assembly and budding are coincident in space and time. Capsid proteins are transported to specific sites on the membrane, where they assemble forming a close structure that is wrapped by the bilayer as it grows (Fig. 1.25). Association of particles in the cytoplasm is rare (although dimers or trimers can form), and assembly appears to be mediated by the binding to the membrane. In some cases, viral glycoproteins not only provide attachment to the cell membrane, but also mediate capsid assembly [63]. Specific segments of viral proteins mediate the ordered association of the capsomers with one another

and are required for proper assembly. Removal of certain sequences results in the assembly of misshapen particles.

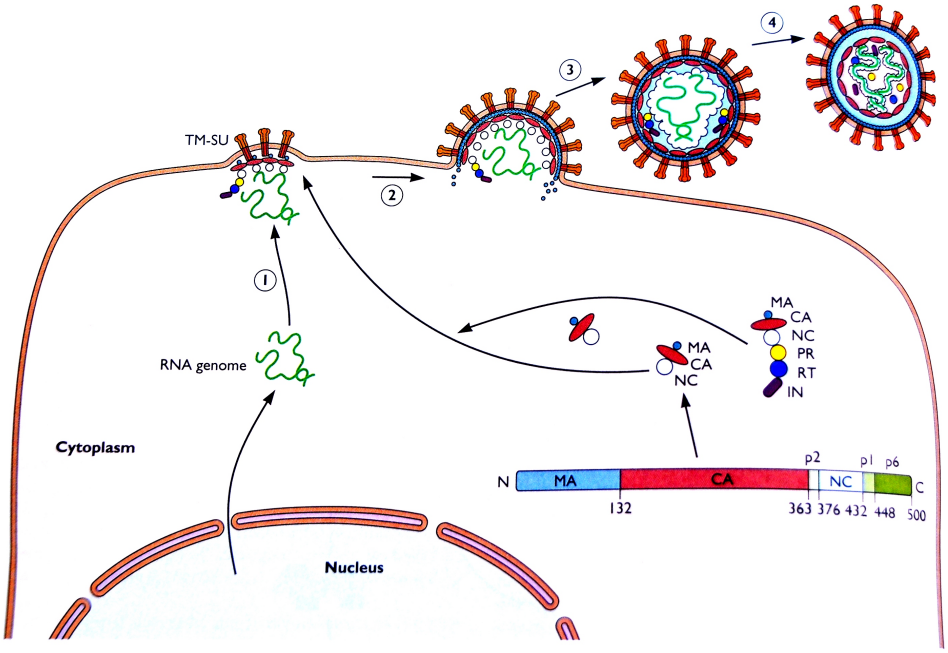

Figure 1.25: Simultaneous assembly from polyprotein precursors and budding. The Gag polyprotein of all retroviruses contains the MA, CA and NC proteins. The association of Gag proteins with the plasma membrane and with the RNA genome initiates assembly at the inner surface of the plasma membrane. Figure adapted from [47]

HIV virus follows this path for assembly and budding, and many of its features have been deeply studied and characterized. Structures obtained by electron microscopy show that the capsid proteins are arranged forming an incomplete sphere in the released virion, the capsomers in fact just cover between 40 and 60% of the inner membrane surface [64, 65]. A completly assembled particle is therefore not necessary for budding to occur, and in some cases, during maturation after budding, the structure is rearranged and the full capsid is formed.

1.3.3 Budding sites

Viral envelopes often present different compositions than the plasma membrane they bud from, which has led to the proposal that viruses bud from specific membrane domains. Multiple evidences in several virus families support the theory that many viruses use lipid rafts as platforms for the entry, assembly and budding [42].

The use of lipid rafts as budding sites can enhance assembly and budding in various ways [66]. Assembly in lipid rafts reduces the effective area where capsomers diffuse before encountering each other, thus capsomer-capsomer interaction is promoted and assembly may

occur in shorter times. Membrane properties of lipid rafts may also improve the characteristics of the budding sites: smaller bending coefficients in those areas may facilitate membrane bending and line tensions may promote budding. Viral proteins might have evolved to bind to characteristic components of the membrane rafts, such as cholesterol and sphingolipids, in order to exploit those possibilities.

In some cases, viruses also take advantage of areas of high curvature to facilitate budding. HIV particles have been seen to bud from deep invaginations localized in the cell membrane of macrophages [67]

1.3.4 Budding driving force

Vesicle generation is an energy cost process due to the high curvature induced in the membrane. The bending energy of an spherical vesicle is $E = 8\pi k_B T = 250 - 600 k_B T$ independently of the radius, for a membrane with bending coefficient $\kappa = 10 - 25 k_B T$. Without taking into account the cost of the complimentary actions during vesicle generation by budding, such as the pore opening or the neck fusion, the high bending energy compared to the thermal energy implies that budding cannot occur spontaneously and needs to be driven.

Budding formation is mainly driven by the energetically favorable self assembly of the capsid, that is coupled to the membrane dynamics via the membrane glycoproteins or the binding matrix. Although the formation of fully infectious particles requires the presence of all components, bud formation is sometimes possible with very few ingredients [68]. For example, matrix protein of vesicular stomatitis virus alone is capable of inducing membrane buds in vitro [69], while expression of coronavirus or flavivirus glycoproteins without any other viral component leads to the release of enveloped particles morphologically similar to complete virions [70]. For HIV and some other retroviruses, Gag alone is sufficient for formation of budded particles. In these cases, it is clear that the protein network does not only attach the capsid to the membrane, but also induces the bud formation. In fact, the matrix proteins also selfassemble in a spherical cage in a similar way than the capsomers do. The viral spikes, in turn, are generally arranged in an icosahedral way [71], so the lateral interactions between them may also be responsible of budding [63].

Membrane properties can also decrease the energy barrier towards budding. High line tension in microdomain borders can induce the protrusion of the membrane out of its plane [37], resulting in the enhancement of curvature in the budding site. Moreover, the concentration of lipids with different intrinsic curvatures in microdomains can also create a spontaneous curvature in the site.

Finally, it is also important to identify the action of the cell machinery during budding. In most cases the ESCRT cellular complex is recruited in the last steps of budding to promote the scission of the bud. Although most enveloped virus use ESCRT, some viruses can carry out both budding and scission without it [72, 73]. Even in the case of HIV, which strongly

depends on ESCRT, particle release has been observed in some types of cells without the ESCRT [74, 75]. On the other hand, also in HIV particles, the action of the machinery complex is also related to an early budding of unfinished capsids [64]. Those dependencies may point that the main role of ESCRT is to speed up a budding process that can occur passively without the cell machinery.

2 Methods

2.1 Introduction

First insights into cell structure arrived with the development of optical microscopy, many organelles were discovered and described with regular light microscopes during the nineteenth century. However, the limitations in their resolution, of around hundreds of nanometers, did not allow the observation of smaller organelles and macromolecules. The study of the structure of the cell membrane or that of viruses was only possible with the development of electron microscopy (Fig. 2.1). With a resolution between 50 and 75 Å, the first viruses and the bilayer structure of the plasma membrane were observed [76]. Cryo-electron microscopy improved the resolution to ranges between 10 to 20Å. Since the diameter of α-helix in proteins is around 10 Å, cryo-electron microscopy used together with three dimensional image reconstruction allows the visualization of virus particles at high detail [77]. Original x-ray microscopy, with resolutions between hundreds of nm to Å, has also been used in the study of membrane and virus structure, provided that the samples yield crystals suitable for x-ray diffraction, which is not always the case. Although the crystallization process may seem harsh in the study of living matter, and it cannot be used in the sample original medium, great discoveries have been done using this technique. X-ray revealed the structure and function of many biological molecules, including all kinds of proteins and DNA. The tobacco mosaic virus was the first to be crystallized [78] and in 1954 its structure was obtained in detail by Watson [79] and Franklin [80]. In the 1980s new techniques of soft X-ray microscopy were developed in a spectral window where water is nearly an order of magnitude more transparent than organics, so that biological specimen can be seen wet and intact [81]

These microscopic techniques give mostly structural information, which is necessary to understand the composition and properties of the systems under study, but it is not enough to understand the dynamic processes ubiquitous in life sciences.

The discovery of the green fluorescence protein and its application to fluorescence microscopy [82] allowed the study of dynamic processes in vivo by tracking fluorescence-labeled molecules. One of the key experiments that validated the general hypothesis of the Fluid Mosaic Model used fluorescence microscopy to track proteins on the lipid membrane and showed their diffusion for the fist time [83]. Besides the determination of diffusion coefficients of lipids and proteins on cellular membranes, fluorescence has been useful in membrane science to identify and characterize the presence of microdomains and lipid rafts

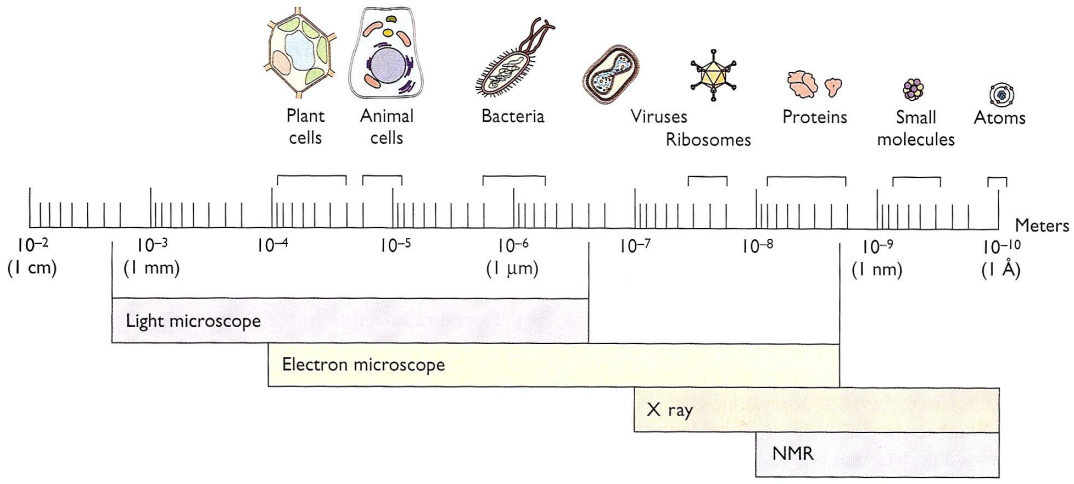

Figure 2.1: Size diversity in biological systems. The size of animal and plant cells, bacteria, viruses, proteins, molecules and atoms are indicated along with the resolving power of microscopy techniques, such as light microscopy, electron microscopy, X-ray crystallography, and nuclear magnetic resonance. The size of membranes goes from hundreds of μm in extension to nm in width, while viruses sizes are about hundreds of nm but their components are in the nm-μm scale. Figure adapted from [47].

[84]. In virology, tracking of individual virus particles in vivo allowed the observation of every stage of the viral cycle, their properties, and the specific molecules involved [85].

Apart from the visualization of structures and behaviors of specific processes, experiments rely on the modification of the parts involved in a process and the observation of the effects of such intervention. Genetic engineering has made possible the alteration of natural systems in order to understand the steps involved in a particular reaction by isolating mutants that are unable to carry out that reaction. With recombinant DNA techniques [86], mutations can be introduced anywhere in the genome of most organisms. An entire sequence can be removed to asses the role of the protein it encodes, or just a single gene can be deleted to understand the functions of specific segments of a coding sequence. The role of many viral proteins in assembly and budding has been identified using these genetic techniques.

Experimental techniques evolve continuously to deal with more and more subtle and sophisticated issues, however, there are cases where the information cannot be obtained experimentally because of a wide variety of difficulties: unreachable reduced scales, technical limitations to fine tune a system or to quantify its properties and behaviors, or impossibility to isolate the variables under study from the whole system or to study their effects unrelated to those from other components.

Ideally, analytical theories could fully address these questions, however, complex systems, in general, cannot be solved analytically. The very simple systems for which analytical solutions can be found are used as approximations, and perturbations can be added in order

to get the best description of the real system. This approach can give information about very general features but it is quite poor in the detail description. In this context, computational simulations arise as a bridge between analytical theories and experiments. They offer a powerful tool to get insight into complex processes by controlling the interaction of every single component and measuring the effect it has on the whole system. Fine tunning the different properties in a wide range of conditions, or deleting and inserting new components are simple procedures. Computer simulations also provide mathematical descriptions of the systems under study that can be easily generalized in order to make general theories that expand the understanding of the subject and predict its possible behaviors.

Simplified systems are used in computational simulations to identify basic features and understand general concepts. In membrane research for instance, due to its complexity, it is often necessary to investigate model bilayers composed of one or two lipid species, and with embedded natural proteins, sterols, or artificial peptides, to get fully insight into their properties. Depending on the nature of the system and the level of detail relevant for the processes under study, different simulation techniques are used.

2.2 Computer simulations: From ab initio to coarse grained

Interesting phenomena in soft matter and biology occur at very different scales (Fig. 2.2). From molecular processes in chemical reactions at a quantum level description to macroscopic processes arising from the collective behavior of millions of molecules, the systems that can be addressed with computational simulations lie in a broad range of spatio-temporal scales. In principle, any system could be studied at the maximum detail through ab initio theories [87], where every atom is explicitly represented at electronic level, and the Schrödinger equation is integrated throughout time. Ab initio simulations techniques are well developed and widely used for a number of different problems [88]. However, when dealing with systems with thousands to millions of complex molecules, these methods are computationally prohibitive. In many cases, the properties under study are far from quantum detail, so the atomic and sub-atomic interactions can be averaged into effective classical interactions. Atomistic models with classical potentials represent a good approximation to these systems, while reducing considerable the simulation time.

In atomistic models, the electronic description is eliminated and the effective interactions between atoms are represented by classical potentials, vastly reducing computational time. These interactions include covalent potentials describing chemical bonds (two particle, bond angle potentials and dihedral, torsional potentials), as well as non-covalent potentials between non bonded atoms, such as electrostatic and van der Waals interactions. The general form of these potentials is showed in Eq.2.1, although the specific functions representing each term can vary from one model to another.

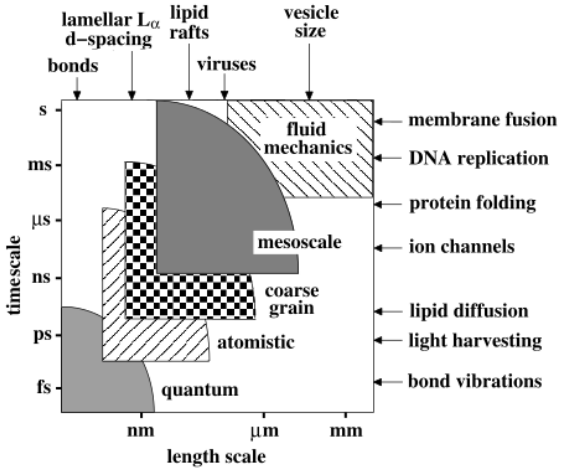

Figure 2.2: Schematic diagram of temporal and spatial scales corresponding with different membrane processes and levels of description. The simulation techniques used to describe the different scales are also depicted. Figure adapted from [89]

$$
\begin{aligned}
V(r^N) &= \sum_{bonds} \frac{1}{2}\kappa_b(l-l_0)^2 + \sum_{angles} \frac{1}{2}\kappa_a(\theta-\theta_0)^2 + \sum_{torsions} \frac{1}{2}V_n[1+\cos(n\omega-\gamma)] + \\
&+ \sum_{j=1}^{N-1}\sum_{i=j+1}^{N} \left\{ \epsilon_{ij}\left[\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{0ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \qquad (2.1)
\end{aligned}
$$

Intramolecular interactions are usually represented through harmonic potentials and are responsible of the molecular geometry, they fix the bond length between atoms, l_0, and the equilibrium angle between groups of 3 atoms, θ_0, or that between planes defined by 4 atoms, γ. A Coulomb potential is usually used for the electrostatic terms and a Lennard-Jones for the van der Waals terms. This type of potential has a short-range repulsive part that accounts for the Pauli repulsion and prevents two atoms from overlapping and an attractive part at long ranges that represents the van der Waals interaction. Although Lennard Jones potential is not the most accurate representation of the van der Waals interaction, due to its computational simplicity, it is used extensively in computer simulations.

These potentials represent qualitatively the type of interactions observed in a general system, but the parameters need to be tuned specifically for every system to describe completely the set of interactions. The whole set of functions and parameters is called the force field. The typical parameter set one needs to specify includes values for atom masses, van der Waals radii, effective partial charges for individual atoms, and equilibrium values of bond lengths, among others. Although theoretically these parameters could be obtained from first-principles calculations, in many particle systems they might be impossible to calculate in practice, and therefore, the parameters are usually fitted using a combination of experi-

mental observables and ab-initio results. Configuration properties like charge and density distribution functions are calculated in ab-initio simulations and directly used to calculate theoretically the potentials between atoms. Other parameters are fitted a posteriori so that the simulations results agree with experiments or ab initio simulations.

Atomistic models manage to describe to some degree the molecular architecture and the interactions between components. In the simulation of lipid bilayers, atomistic simulations are able to correlate the detailed lipid architecture (e.g., particular head group structure, length, and degree of saturation of the tails) with the physical properties of the bilayer (thickness, orientation of segments, liquid-crystalline ordering, and elastic moduli) [90]

Classical potentials reduce considerably computation time, however, atomistic detail is unnecessary for most systems with relevant properties at the mesoscopic scale. Collective phenomena involve many molecules and entail large length and time scales, 10–1000 nm and μs–ms. Although these phenomena arise from the numerous degrees of freedom of the whole system and their microscopic interactions, the meso- or macroscopic observables of interest do not in general depend on all the details of the atomic description, because many microstates map to the same meso- or macrostate [91]. Coarse graining techniques aim to identify, among all the system interactions, the minimum variables capable of describing the specific macroscopic behaviors under study. Two different coarse graining approaches have been developed: systematic and generic coarse graining.

In the first case, the models are derived systematically from the atomistic systems, by integrating out those atomic coordinates that are not relevant for the process under study. In this way, the atoms are grouped together into new effective particles, or beads, that lack of atomistic detail. The effective interactions between the new units define the coarse grained force field. They are obtained in a similar way that for atomistic models, and the same type of generic interactions are often used (eq. 2.1), except that the information to fit the parameters comes from a combination of experimental observables and atomistic simulation results. This strategy has the advantage that detailed structural information is incorporated into the model, but suffers from loss of transferability; since the functions obtained from atomistic simulations and used to fit the parameters incorporate temperature, density, composition, and other dependences into the effective pair interaction, the resulting force field can have a severely limited range of applicability, being valid only for systems under similar conditions than that used to parametrize against [89]. In Fig. 2.3 the coarse graining developed by [92] for a dimyristoylphosphatidylcholine is shown, the CG model consists of 13 sites and 24 internal potentials (12 bonds and 12 bends), while the all-atom model encompasses 118 atoms and 971 internal potentials (117 bonds, 226 bends, 315 torsions, and 313 one-fours) [89]. This significative reduction on the degrees of freedom allows to simulate on larger spatio-temporal scales.

In generic coarse graining however, idealized models that incorporate only essential properties to describe a specific behavior are used (Fig. 2.3). These models are usually less quan-

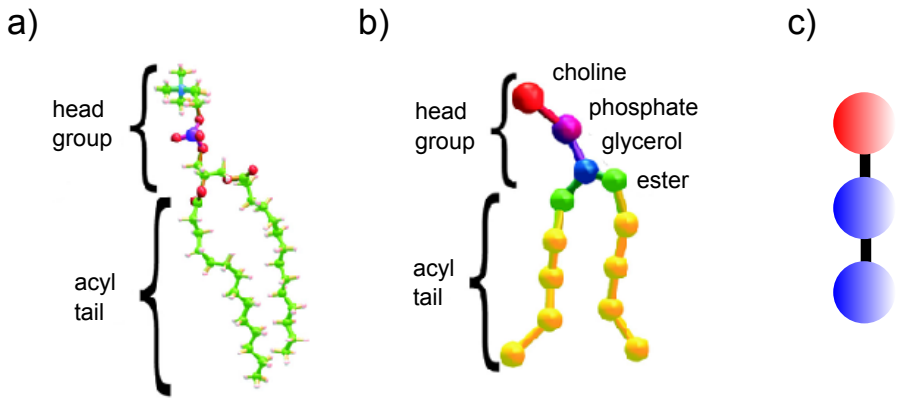

Figure 2.3: Schematic showing the coarse graining process. Groups of atoms are lumped together into coarse grained beads that keep the essential properties of the molecule affecting a specific process. a) Atomistic representation of DMPC. b) Coarse grain representation presented by [92] of DMPC with 13 beads, containing information of the main moieties. c) Coarse grained model of a generic phospholipid used by [93], where only three beads are needed to represent the main features of an amphiphile: the hydrophilic head and the hydrophobic tail

titative but give insight into basic physical mechanisms often unaccessible by other means. The potential form used for the interactions between beads is usually chosen intuitively and the parameters are fitted a posteriori after studying the model properties. Although they can usually account only for the specific properties they have been designed to display, these models can be very powerful in the study of systems with large time and length scales.

Since coarse grained models are designed to be as efficient as possible, the removal of the solvent results to be very appealing to further reduce the computation time. The solvent is an important time-consuming source that represents a substantial fraction of the system. Implicit solvent models eliminate the explicit representation of the solvent and replace it by effective interactions between the remaining solute particles, this reduces substantially the degrees of freedom. Although this is a common practice in most coarse grained polymer systems, its implementation in amphiphile systems can be very tricky, since the solvent we want to remove helps to drive the assembly of amphiphiles. Despite of this, effective forces can be carefully chosen to avoid the use of the solvent. In some cases however, the molecular interactions with the solvent or the hydrodynamics play a crucial role in the mechanism under study and cannot be removed. An effective way to introduce the hydrodynamic effects on these type of models is via dissipative particle dynamics (DPD) [94, 95].

The softer intermolecular potentials on the coarsened scale allow the use of larger time steps when integrating the equations of motion for the particles in the system. This, together with the reduced number of degrees of freedom, lead to a significant computational speed-

up with the consequence that larger systems and longer time scales are accessible. However, these artifacts can lead to systems with distorted dynamics, in fact, the softer interactions typically speed up the dynamics on large length scales [90]. Nevertheless, the system dynamics can be rescaled relative to the natural time scale of the model, that is usually chosen as the characteristic time of the slowest process of relevance.

Although the previous approximations make coarse graining very efficient, the loss of detail and its subsequent effects need to be estimated in oder to understand the limitations of the model. As usual, the application of a model to a system is subjected to its validation, which may consist on investigating its behavior and comparing it to actual results from experiments or atomistic simulations. The typical observables that are used to set the validity of the models are usually related to equilibrium quantities or dynamic coefficients. Depending on the approach to the system, sometimes models do not necessarily need to reproduce exactly the characteristics of the system, but only in a qualitative level. This does not mean that quantitative information cannot be obtained with these models, but the observables will be necessarily related to the system scales.

Despite of the apparent limitations, and thanks to the high efficiency of coarse grained models, many systems of interest have been studied and characterized with this technique. In the study of membranes, for example, those models have been applied to a great diversity of problems, such as the study of the phase behavior of lipid bilayers [96, 97], the effect of phospholipid double [27, 98] or single tails [99, 100], mixed bilayers of different phospholipids [101], effects of cholesterol [102], interactions between proteins and lipid bilayers [103, 104], protein insertion into membranes [105, 106], vesicle formation [107], or membrane fusion [108].

Summarizing, biological systems span a broad range of time and length scales, and different approaches can be chosen to deal with such diversity. Different techniques are used in the study of complex systems, depending on the scale of the processes under study, with the common goal of obtaining the equilibrium states to get a thermodynamic description of the system. In the study of phenomena at macroscopic length and time scales continuum models can be used. In membrane research this approach is appropriate in the analysis of shape deformation and thermal fluctuations of vesicles and lipid bilayers.

However, when the phenomena of interest occur in the mesoscopic time and length scale, or when the effect of the molecular structure of the membrane components on the physical behavior of the whole system is to be elucidated, a particle-based approach becomes necessary [109]. The simulation techniques mostly used to implement these models are molecular dynamics (MD) and Monte Carlo (MC). In MD simulations, the equations of the motion are integrated for every particle in the system throughout time. The full set of positions and momenta defines a trajectory from which the dynamic behavior of the system is obtained. The macroscopic observables are calculated by averaging these quantities over time. In MC simulations however, random configurations are accepted or rejected according to a probability

defined by the system Hamiltonian, until the equilibrium distribution is achieved [110, 111]. MC simulation can be more efficient in finding the steady states, however, dynamic properties can only be inferred from energy comparisons between states or if the Monte Carlo move set can be mapped onto dynamical processes. Since the subject of this work is a dynamic process, MD simulations will be the main tool and the following sections will address their most important characteristics.

2.3 Molecular Dynamics simulations

In a MD simulation, given a Hamiltonian, the equations of motion of every particle in a system are integrated to calculate their trajectories. The Hamiltonian, $\mathcal{H}(\mathbf{q}^N, \mathbf{p}^N)$, of a general system expresses the total energy as a function of the generalized positions and momenta: $\mathcal{H} = T + \mathcal{U}$, where T is the kinetic energy of the system and $\mathcal{U}$ is the potential energy. In a classical system of interacting particles, the potential energy has the general form presented in Eq. 2.1. In Cartesian coordinates, Hamilton's equation of motion become a set of second order differential equations:

$$m_i \ddot{\boldsymbol{r}}_i = \boldsymbol{f}_i \tag{2.2}$$

where $\boldsymbol{r}_i$ and m_i are the position and mass of a particle i, and $\boldsymbol{f}_i$ is the force exerted on particle i by the rest of the system. The equations of motion can also be represented by a set of two first order differential equations:

$$\dot{\boldsymbol{r}}_i = \boldsymbol{p}_i / m_i \tag{2.3}$$

$$\dot{\boldsymbol{p}}_i = -\nabla_{\boldsymbol{r}_i} \mathcal{U} = \boldsymbol{f}_i \tag{2.4}$$

where $\boldsymbol{p}_i$ is the momentum of particle i.

Therefore, computing the trajectory of every particle involves solving either a set of 3N second order differential equations (eq. 2.2) or an equivalent set of 6N first order differential equations (eqs. 2.3 and 2.4). In most cases, the classical Hamiltonians are time-independent, which leads to energy conservation, an important feature that a decent algorithm should ensure.

2.3.1 Algorithms

A standard method for the integration of ordinary differential equations such as eqs. 2.3 and 2.4 is the finite difference approach. This method consists on obtaining the particle positions and velocities at certain time $t + \Delta t$, given the system variables at a previous time t. The equations are solved on a step by step basis and the choice of the time interval will depend

on the method of resolution and the system properties, but it has to be considerably smaller than the shortest characteristic time in a system, e.g., the vibrational timescale for covalent bonds involving hydrogens in atomistic simulations.

In general, a good algorithm should be fast, efficient, accurate and simple. In MD simulations the speed of the integration method is not crucial because the fraction of time used to integrate the equations of motion is always very small compared to the time needed to calculate the interactions between particles. Efficiency can be understood in different ways, on one hand the minimization of the resource usage, for example by storing the smallest possible number of variables, will require less computer memory and will allow the study of bigger systems. On the other hand, an algorithm is also efficient when it spans longer times with less calculations. This can be achieved setting long time steps Δt, so a given period of simulation will cover longer system times. This feature can reduce noticeably the simulation time, however, longer time steps are also related with worse accuracy, which becomes apparent at large times when the system trajectory diverges from the correct classical trajectory. Any two classical trajectories which are initially very close will diverge exponentially with time with a speed characterized by a Lyapunov exponent [112]. Any tiny error associated with finite precision arithmetic will tend to cause these divergences. Although no integration algorithm will provide and exact solution for long times, a good one will generate trajectories close enough to the the true values.

The inability to exactly generate real trajectories is not a drawback, since the aim of MD is not to specify the exact microstates of the system at every time. In MD simulations the magnitudes of interest are obtained from statistical samples, so a trajectory close enough to the real one will be equally valid provided the constants of motion are conserved. Algorithms not showing energy conservation will not generate correct ensemble averages, so the statistical predictions obtained from those simulations will be wrong. Energy conservation is degraded as time step is increased, because any finite difference algorithm is only exact for the limit where time steps tend to zero. Therefore, there will be a trade-off in the time step length between efficiency and accuracy. A good algorithm acceptably preserves energy conservation for relatively large time steps. Other important factors related to energy conservation are the shape of the potential and the typical particle velocities. Thus, shorter time steps are used at high temperatures, for light molecules, and for rapidly varying potential functions in order to ensure well behaving simulations [110].

One of the most simple and extended method to solve eqs. 2.3 and 2.4 is the **Verlet algorithm**. It is built from a Taylor expansion of the position around time t using eq. 2.2:

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\Delta t + \frac{\boldsymbol{f}(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\dddot{\boldsymbol{r}} + \mathcal{O}(\Delta t^4) \tag{2.5}$$

similarly,

$$\boldsymbol{r}(t-\Delta t) = \boldsymbol{r}(t) - \boldsymbol{v}(t)\Delta t + \frac{\boldsymbol{f}(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}\dddot{\boldsymbol{r}} + \mathcal{O}(\Delta t^4) \tag{2.6}$$

and summing eqs. 2.5 and 2.6 we obtain

$$\boldsymbol{r}(t+\Delta t) + \boldsymbol{r}(t-\Delta t) = 2\boldsymbol{r}(t) + \frac{\boldsymbol{f}(t)}{m}\Delta t^2 + \mathcal{O}(\Delta t^4) \tag{2.7}$$

or equivalently,

$$\boldsymbol{r}(t+\Delta t) \approx 2\boldsymbol{r}(t) - \boldsymbol{r}(t-\Delta t) + \frac{\boldsymbol{f}(t)}{m}\Delta t^2 \tag{2.8}$$

Here, Δt is the time step in the Verlet algorithm and the error of the estimation of the coordinates at time $t + \Delta t$ is of order Δt^4. As it is clear from eq. 2.8, the Verlet algorithm does not use the velocity to compute the positions, however, velocity is usually needed for the calculation of the kinetic energy or of other magnitudes, and it is easily obtained from the positions along the trajectory:

$$\boldsymbol{v}(t) = \frac{\boldsymbol{r}(t+\Delta t) - \boldsymbol{r}(t-\Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \tag{2.9}$$

The velocity is only accurate to order Δt^2. The Verlet algorithm is relatively fast, it requires little memory and it exhibits little long-term energy drift [111]. Among its limitations, it is not particularly accurate for long time steps, the velocities are handled in an awkward way and it presents numerical imprecision. The numerical imprecisions are due to the small term ($\mathcal{O}(\Delta t^2)$) that is added to a difference of large terms ($\mathcal{O}(\Delta t^0)$) in eq. 2.8 to find the new positions.

Some modifications have been developed to reduce the deficiencies of the Verlet algorithm, all of them produce identical trajectories, the differences between them are more for convenient score precision in certain averages. A simple one is the **Leap-frog algorithm**, which is very similar to the Verlet scheme but evaluates the velocities at half-integer time steps and uses them to compute the new positions.

$$\boldsymbol{v}(t+\Delta t/2) = \boldsymbol{v}(t-\Delta t/2) + \frac{\boldsymbol{f}(t)}{m}\Delta t \tag{2.10}$$

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t+\Delta t/2)\Delta t \tag{2.11}$$

Although the velocities appear explicitly in the Leap-frog scheme, they still need to be calculated at the mid-step values, so that the kinetic and potential energies are evaluated at the same time to obtain the total energy. Even so, there are still precision benefits from the fact that the difference between two large quantities are not used to get a small one. There is another algorithm, mostly preferred nowadays, that deals in a more natural way with the velocities and reduces the precision problems. This **two step velocity Verlet algorithm** is

similar to the original one but differs in the fact that the positions and velocities are updated in two steps:

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\Delta t + \frac{\boldsymbol{f}(t)}{2m}\Delta t^2 \tag{2.12}$$

$$\boldsymbol{v}(t+\Delta t) = \boldsymbol{v}(t) + \frac{\boldsymbol{f}(t)+\boldsymbol{f}(t+\Delta t)}{2m}\Delta t \tag{2.13}$$

This algorithm involves two stages with a force evaluation in between. Firstly, the new positions at time $t+\Delta t$ are calculated using eq. 2.12, and the velocities at mid-time step are evaluated using:

$$\boldsymbol{v}(t+\Delta t/2) = \boldsymbol{v}(t) + \frac{\boldsymbol{f}(t)}{2m}\Delta t \tag{2.14}$$

The forces at time $t+\Delta t$ are then computed with the new positions, and the velocity update completed:

$$\boldsymbol{v}(t+\Delta t) = \boldsymbol{v}(t+\Delta t/2) + \frac{\boldsymbol{f}(t+\Delta t)}{2m}\Delta t \tag{2.15}$$

The velocities and positions are obtained at the end of each step, so the kinetic and potential energies can be easily evaluated at the same time.

Other algorithms, such as the Gear predictor-corrector, are also used to integrate the equations of motion in MD [110].

2.3.2 Rigid non-spherical bodies

The algorithms presented in the previous section can only be applied to point-like or spherical symmetric particles. In atomistic models, particles belonging to the same molecule interact via intramolecular potentials which are around one order of magnitude stronger than intermolecular potentials. Therefore, in some cases, and specially in coarse grained models, some intramolecular bonds are fixed, so the number of degrees of freedom is reduced and the molecule is treated as a rigid body.

A rigid body b is composed of N_b particles , its center of mass is located at $\boldsymbol{R}_b$ and moves with velocity $\boldsymbol{V}_b$, its total mass is M_b and its moment of inertia and angular momentum are $\boldsymbol{I}_b$ and $\boldsymbol{L}_b$ respectively. The orientation of the body specifies the relation between an axis system fixed in space and another fixed with respect to the body, usually this is the 'principal' one where the body inertia tensor is diagonal. Any unit vector $\boldsymbol{e}$ may be expressed related to the body-fixed frame, $\boldsymbol{e}^B$, or the space-fixed frame, $\boldsymbol{e}^s$. Both are related by the rotation matrix $\boldsymbol{A}$

$$\boldsymbol{e}^B = \boldsymbol{A}\cdot\boldsymbol{e}^s \tag{2.16}$$

The classical Euler description for the rotation of the rigid body defines any rotation in the three dimensional space with the use of three generalized angles and the following rotation matrix:

$$\boldsymbol{A} = \begin{pmatrix} \cos\phi\cos\psi - \sin\phi\cos\theta\sin\psi & \sin\phi\cos\psi + \cos\phi\cos\theta\sin\psi & \sin\theta\sin\psi \\ -\cos\phi\sin\psi - \sin\phi\cos\theta\cos\psi & -\sin\phi\sin\psi + \cos\phi\cos\theta\cos\psi & \sin\theta\cos\psi \\ \sin\phi\sin\theta & -\cos\phi\sin\theta & \cos\theta \end{pmatrix} \tag{2.17}$$

The equations of motion for the Euler angles are given by [113]:

$$\begin{aligned} \dot{\phi} &= -\omega_x^s \frac{\sin\phi\cos\theta}{sin\theta} + \omega_y^s \frac{\cos\phi\cos\theta}{\sin\theta} + \omega_z^s \\ \dot{\theta} &= -\omega_x^s \cos\phi + \omega_y^s \sin\phi \\ \dot{\psi} &= -\omega_x^s \frac{\sin\phi}{\sin\theta} + \omega_y^s \frac{\cos\phi}{\sin\theta} \end{aligned} \tag{2.18}$$

where $\boldsymbol{\omega}^s$ is the angular frequency in the space-fixed frame.

The presence of $\sin\theta$ in the denominator of various terms in eq. 2.18 means that a divergence occurs when θ approaches 0 or π. This singularities do not permit the formulation of a numerical stable set of equations. Other schemes have been proposed to deal with this instability, such as the redefinition of the Euler angles when a singularity is approached [114], but the most simple and widespread algorithm is based on the use of quaternions. With the inclusion of an extra generalized coordinate, the equations of motion can be rewritten for the new system variables so they are well defined and present no singularities. The quaternions used in rigid body rotation are four dimensional vectors, $\boldsymbol{Q} = (q_0, q_1, q_2, q_3)$, satisfying $\sum_i q_i^2 = 1$, that represent the orientation of the rigid body.

They can be defined as functions of Euler angles:

$$\begin{aligned} q_0 &= \cos\frac{\theta}{2}\cos\frac{\phi+\psi}{2} \\ q_1 &= \sin\frac{\theta}{2}\cos\frac{\phi-\psi}{2} \\ q_2 &= \sin\frac{\theta}{2}\sin\frac{\phi-\psi}{2} \\ q_3 &= \cos\frac{\theta}{2}\sin\frac{\phi+\psi}{2} \end{aligned} \tag{2.19}$$

so the rotation matrix in the quaternion formulation becomes

$$\boldsymbol{A}(\boldsymbol{Q}) = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 + q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 - q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 + q_0q_1) \\ 2(q_1q_3 + q_0q_2) & 2(q_2q_3 - q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} \tag{2.20}$$

which is orthogonal, i.e. $\boldsymbol{A}^{-1} = \boldsymbol{A}^T$. The quaternion satisfies the equation of motion $\dot{\boldsymbol{Q}} = \frac{1}{2}\boldsymbol{S}(\boldsymbol{Q})\boldsymbol{\Omega}^B$, or equivalently:

$$\begin{pmatrix} \dot{q}_0 \\ \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \begin{pmatrix} 0 \\ \omega_x^B \\ \omega_y^B \\ \omega_z^B \end{pmatrix} \tag{2.21}$$

where $\boldsymbol{\Omega}^B$ is a four dimensional vector derived from the angular momentum in the body-fixed frame.

The equations of motion are thus defined by $\boldsymbol{S}(\boldsymbol{Q})$, which represents a system of first-order differential equations without singularities. In the quaternion formulation the position and velocity in the space frame of the particle k belonging to the body b are given by:

$$\begin{aligned} \boldsymbol{r}_{bk} &= \boldsymbol{R}_b + \boldsymbol{A}(\boldsymbol{Q}_b) \cdot \boldsymbol{D}_{bk} \\ \boldsymbol{v}_{bk} &= \boldsymbol{V}_b + \boldsymbol{\omega}_b \times (\boldsymbol{A}(\boldsymbol{Q}_b) \cdot \boldsymbol{D}_{bk}) \end{aligned} \tag{2.22}$$

where $\boldsymbol{D}_{bk}$ is a vector that defines the position of the particle relative to the center of mass in the body frame, and $\boldsymbol{\omega}_b$ is the body angular velocity about is center of mass in the space frame, $\boldsymbol{\omega}_b = \boldsymbol{A}(\boldsymbol{Q}_b)\boldsymbol{I}_b^{-1}\boldsymbol{A}^T(\boldsymbol{Q}_b)\boldsymbol{L}_b$. Notice that the angular momentum in the space frame, $\boldsymbol{L}_b$, is transformed to the body frame, so the inertia tensor is diagonal and $\boldsymbol{I}^{-1}$ is directly obtained.

The net force $\boldsymbol{F}$ and torque $\boldsymbol{\tau}$ acting on body b in the space frame are the sum of the individual forces and torques on each particle of the body resulting from the interactions between particles belonging to different rigid bodies.

$$\begin{aligned} \boldsymbol{F}_b &= \sum_{k=1}^{N_b} \boldsymbol{f}_{bk} \\ \boldsymbol{\tau}_b &= \sum_{k=1}^{N_b} [\boldsymbol{A}(\boldsymbol{Q}_b) \cdot \boldsymbol{D}_{bk}] \times \boldsymbol{f}_{bk} \end{aligned} \tag{2.23}$$

Finally, the equations of motion for the translational and rotational coordinates of the rigid body are:

$$\begin{aligned} \dot{\boldsymbol{R}}_b = \boldsymbol{V}_b \quad , \qquad \dot{\boldsymbol{V}}_b = \boldsymbol{F}_b / M_b \\ \dot{\boldsymbol{L}}_b = \boldsymbol{\tau}_b \quad , \qquad \dot{\boldsymbol{Q}}_b = \tfrac{1}{2}\boldsymbol{S}(\boldsymbol{Q}_b)\boldsymbol{\Omega}_b \end{aligned} \tag{2.24}$$

The simple Leap-frog algorithm cannot be directly applied to these equations of motion because the quaternion derivatives $\dot{\boldsymbol{Q}}$ depend not only upon the angular velocity but also on $\boldsymbol{Q}$ itself. Nevertheless this limitation can be overcome with a modified two-step Leap-frog

algorithm where quaternions are moved half step after the update of the angular momentum from which the angular velocity is obtained [115]. The quaternions, $\boldsymbol{Q}(t)$, and torques $\boldsymbol{\tau}(t)$ are obtained from the positions and orientations at time t, and from them, the angular momenta are updated:

$$\boldsymbol{L}(t) = \boldsymbol{L}(t - \Delta t/2) + \frac{1}{2}\boldsymbol{\tau}(t)\Delta t \tag{2.25}$$

The angular momentum is used to obtain the angular velocity, $\boldsymbol{\omega}(t)$, at time t, so the time derivative of the quaternion, $\dot{\boldsymbol{Q}}(t)$ can be calculated from eq. 2.21. Then $\boldsymbol{Q}(t + \Delta t/2)$ is updated at half time step:

$$\boldsymbol{Q}(t + \Delta t/2) = \boldsymbol{Q}(t) + \frac{1}{2}\dot{\boldsymbol{Q}}(t)\Delta t \tag{2.26}$$

The same process is repeated to obtain $\boldsymbol{Q}(t)$.

Both algorithms for rotational and translational movement can be combined, so that the dynamics are obtained at the same time. The velocity and angular momentum of each rigid body are first updated to half step, and the position and orientation are moved a full step by the equations:

$$\begin{aligned}
\boldsymbol{V}(t + \Delta t/2) &= \boldsymbol{V}(t) + \frac{\boldsymbol{F}(t)}{2M}\Delta t \\
\boldsymbol{R}(t + \Delta t) &= \boldsymbol{R}(t) + \boldsymbol{V}(t + \Delta t/2)\Delta t \\
\boldsymbol{L}(t + \Delta t/2) &= \boldsymbol{L}(t) + \boldsymbol{\tau}(t)\frac{\Delta t}{2} \\
\boldsymbol{Q}(t + \Delta t) &= \boldsymbol{Q}(t) + \frac{1}{2}\dot{\boldsymbol{Q}}(t + \Delta t/2)\Delta t
\end{aligned} \tag{2.27}$$

Forces and torques are then calculated based on the updated positions and orientations, and the velocity and angular momentum are advanced a second half step:

$$\begin{aligned}
\boldsymbol{V}(t + \Delta t) &= \boldsymbol{V}(t + \Delta t/2) + \frac{\boldsymbol{F}(t + \Delta t)}{2M}\Delta t \\
\boldsymbol{L}(t + \Delta t) &= \boldsymbol{L}(t + \Delta t/2) + \boldsymbol{\tau}(t + \Delta t)\frac{\Delta t}{2}
\end{aligned} \tag{2.28}$$

2.3.3 Simulating in different Ensembles

Statistical mechanics defines the expected value of an observable as the average of its related magnitude over all the possible states in the system. However, MD simulations are used to obtain macroscopic observables in many particle systems by taking their averages over time. Thus, MD approach is only valid when the ergodic hypothesis holds true, i.e., time averages coincide with ensemble averages. For this to happen, the dynamics of the system has to

ensure that all the microstates are accessible according to their probability. The microcanonical ensemble (NVE) represents an isolated system composed of N particles in a volume V where the energy is a constant of the motion. Since there is no energy or particle exchange, the total energy does not fluctuate and all the accesible microstates have the same energy and are equally likely to be occupied. Since the classical equations of motion conserve the total energy, all the states with energy E are accesible during the system evolution, and so, this ensemble is ergodic with respect to Newton dynamics.

However, biological systems are not isolated, and therefore correspond to constant temperature (NVT) or constant pressure ensembles (NPT) rather than the microcanonical ensemble. Although Newton's equations of motion generate only trajectories at constant energy, it is possible to find NVT and NPT microstates for any energy, that cannot be linked by classical trajectories. In those cases, MD can still be used with modified equations of motion, that do not necessarily have any physical interpretation but generate a succession of configurations in accordance with the correct probability density, so the time average equals the ensemble average, and the correct equilibrium properties can be calculated. The modified dynamics in principle do not reproduce the real one, although deviations of the motion from the truly Newtonian trajectory may be very small.

Canonical (NVT)

Several methods have been developed to extend MD for systems at constant temperature by coupling the system to a heat bath, but not all of them generate the correct canonical ensemble and reproduce the real dynamical properties. The two main approaches to keep temperature constant are based on the rescaling of the velocities so that the kinetic energy is controlled, or the formulation of an extended Lagrangian with new degrees of freedom that represent the dynamics of the thermal bath.

In the **Andersen thermostat**, the coupling between the system and the heat bath is represented by stochastic forces that act at a certain frequency on randomly selected particles. Between those statistical collisions, the system evolves in the phase space on a surface of constant energy according to Newton's laws, and changes from one surface to other due to energy gains and losses during collisions occur according to the Boltzmann probability. This method generates the canonical distribution but is not suitable for the study of dynamical properties. The stochastic collisions disturb the dynamics and lead to a decay of the velocity correlation function which affects the diffusion coefficient at high collision frequencies. For small collision rates, the change in the diffusion coefficient can be very small, however, the dynamics generated by the Andersen thermostat are unphysical [111].

Another approach is followed in the **Nosé-Hoover thermostat** to obtain deterministic MD at constant temperature. A new degree of freedom for the reservoir dynamics is added to the Lagrangian, eq. 2.29, and the simulation is carried out for the extended system.

$$\mathcal{L}_{NOSE} = \sum_{i=1}^{N} \frac{m_i}{2} \xi^2 \dot{\boldsymbol{r}}_i^2 - \mathcal{U}(\boldsymbol{r}^N) + \frac{Q}{2}\dot{\xi}^2 - g k_{\mathrm{B}} T \ln\xi \tag{2.29}$$

The extra degree of freedom for the reservoir is ξ and its conjugate momentum p_ξ, and g is the number of independent degrees of freedom. Energy flows from the reservoir to the system and back with a certain inertia regulated by Q, which also determines the temperature fluctuations. High values of Q result in low energy flow between the system and the reservoir, for Q tending to infinity the conventional MD is retrieved. On the other hand, low values of Q lead to weak, damped oscillations of energy, resulting in poor equilibration. For a specific system, Q is usually chosen by trial and error until satisfying oscillations are achieved.

From the extended Lagrangian, the equations of motion for the system and the reservoir are obtained:

$$\begin{aligned}
\dot{\boldsymbol{r}}_i &= \frac{\boldsymbol{p}_i}{m_i} \\
\dot{\boldsymbol{p}}_i &= -\frac{\partial \mathcal{U}(\boldsymbol{r}^N)}{\partial \boldsymbol{r}_i} - \gamma \boldsymbol{p}_i \\
\dot{\gamma} &= \left(\sum_i \frac{\boldsymbol{p}_i^2}{m_i} - \frac{L}{\beta} \right) / Q \\
\frac{\dot{\xi}}{\xi} &= \frac{\mathrm{d}\ln\xi}{\mathrm{d}t} = \gamma
\end{aligned} \tag{2.30}$$

where $\gamma = \xi p_\xi / Q$ is the reservoir friction coefficient.

Many other different ways to sample the canonical ensemble have been proposed, but that of **Langevin dynamics** (LD) is of special interest for the type of systems analyzed in this work. The Langevin equation, eq. 2.31, describes the dynamics of many particles system, where the collisions between simulated and implicit particles are represented by stochastic and friction forces:

$$m\ddot{\boldsymbol{r}} = -\boldsymbol{\nabla}\mathcal{U}(\boldsymbol{r}) - \gamma m \dot{\boldsymbol{r}} + \sqrt{6\gamma k_B T m}\boldsymbol{R}(t) \tag{2.31}$$

where γ is the friction coefficient, T is the temperature of the system, and $R(t)$ is a delta-correlated stationary Gaussian process with zero mean:

$$< R(t) >= 0 \quad , \quad < R(t)R(t') >= \delta(t - t') \tag{2.32}$$

The average magnitude of the random forces and the friction are related so that the fluctuation-dissipation theorem is satisfied and therefore, the temperature is kept constant, so that a correct sample of the canonical ensemble is obtained. γ is related to the diffusion coefficient by the fluctuation-dissipation theorem, as it grows, the dynamics change from the

inertial to the diffusive regime where no average accelerations take place, this is described by Brownian Dynamics [116]

$$\dot{\boldsymbol{r}} = -\frac{\boldsymbol{\nabla}\mathcal{U}(\boldsymbol{r})}{\zeta} + \sqrt{6D}\boldsymbol{R}(t) \tag{2.33}$$

where $\zeta = \gamma m$ is the drag coefficient and $D = k_B T/\zeta$ the diffusion coefficient.

Langevin and Brownian dynamics are widely used in implicit solvent models to keep the Brownian motion of a particle in water without representing it explicitly. Another advantage of the Langevin thermostat, and in general, of stochastic thermostats, is that simulations are stable using larger time steps.

Langevin Dynamics is therefore an efficient algorithm with a twofold ability: it gives an approximate representation of Brownian dynamics while sampling the canonical ensemble.

Isobaric (NPT)

It was in the development of a barostat, when Andersen proposed for the first time that an extended system could be used to apply MD in other ensembles [117]. In the isobaric case, the pressure is kept constant by changing the volume, the new dynamical variable that couples the system to the barostat, mimicking the action of a piston on a real system. The extended Lagrangian is written for the scaled positions, $\boldsymbol{s}$, and velocities $\dot{\boldsymbol{s}}$, referred to the instantaneous volume:

$$\boldsymbol{r} = V^{1/3}\boldsymbol{s} \quad , \quad \boldsymbol{v} = V^{1/3}\dot{\boldsymbol{s}} \tag{2.34}$$

so the extended Lagrangian becomes

$$\mathcal{L} = \sum_i \frac{V^{2/3}}{2} m_i \dot{\boldsymbol{s}}_i^2 - \sum_{i<j} \mathcal{U}_{ij}(V, \{\boldsymbol{s}_i\}) + \frac{W}{2}\dot{V}^2 - PV \tag{2.35}$$

where W is the piston mass and is related to the piston inertia. A low value of W results in rapid box size oscillations which are not damped very efficiently . A large value however, results in a slow exploration of the volume-space, while an infinite W restores the normal MD. In homogeneous systems, the changes in volume in the simulation box are isotropic and are applied equally in the three lateral dimensions. For inhomogeneous systems, more complex algorithms are used where not only the volume of the simulation box changes to keep the pressure constant, but also its shape [118]. The equations of motion obtained from the Lagrangian in eq. 2.35 generate trajectories which sample the isobaric ensemble.

In many cases the barostat is coupled to a thermostat of the types presented in the previous sections in order to simulate in the isothermal-isobaric ensemble (NPT). Martyna [119] proposed the equations of motion for the extended Lagrangian that couples the Nosé-Hoover thermostat with an Andersen-like barostat that generates the correct distribution in the NPT ensemble for a d-dimensional system of N particles with N_f =Nd degrees of freedom:

$$\dot{\boldsymbol{r}} = \frac{\boldsymbol{p}}{m} + \frac{p_\epsilon}{W}\boldsymbol{r}, \qquad \dot{\boldsymbol{p}} = -\nabla\mathcal{U} - \left(1 + \frac{d}{N_f}\frac{p_\epsilon}{W}\boldsymbol{p} - \frac{p_\xi}{Q}\boldsymbol{p}\right) \tag{2.36}$$

$$\dot{V} = \frac{dVp_\epsilon}{W}, \qquad \dot{p}_\epsilon = dV(\mathcal{P} - P_0) + \frac{d}{N_f}\sum_{i=1}^{N}\frac{\dot{\boldsymbol{p}}_i{}^2}{m_i} - \frac{p_\xi}{Q}p_\epsilon \tag{2.37}$$

$$\dot{\xi} = \frac{p_\xi}{Q}, \qquad \dot{p}_\xi = \sum_{i=1}^{N}\frac{\dot{\boldsymbol{p}}_i^2}{m_i} + \frac{p_\epsilon^2}{W} - (N_f+1)k_BT \tag{2.38}$$

where $\boldsymbol{r}$ and $\boldsymbol{p}$ are the position and momentum of a particle, V is the volume, p_ϵ is the barostat momentum, ξ and p_ξ are the thermostat position and momentum, $\mathcal{P}$ and P_0 are the instantaneous and target pressures, and Q and W are the artificial piston masses for the thermostat and the barostat respectively. The instantaneous pressure is obtained from the virial and kinetic terms, and for pairwise potentials takes the form:

$$\mathcal{P} = \frac{1}{\mathrm{d}V}\sum_{i<j}\boldsymbol{f}_{ij}\boldsymbol{r}_{ij} + \frac{1}{\mathrm{d}V}\sum_i\frac{1}{m_i}\boldsymbol{p}_i^2 \tag{2.39}$$

These equations of motion can be integrated using a Verlet algorithm, where the volume at a time t is updated using the volume conjugate momentum in the previous half-step, which is related to the difference between the instantaneous and target pressures. Then, the positions and velocities are rescaled, so they adapt to the changes in volume.

In part of the work presented here, the Verlet scheme proposed by Kolb et al [120] to sample the NPT ensemble will be adapted to the specific features of the systems under study. This scheme couples a Langevin thermostat with an Anderson barostat using Π_V, the conjugate momenta of V, with $\dot{V} = \Pi_V/W$, and proceeds as follow:

1. $\boldsymbol{p}_i' = \boldsymbol{p}_i(t) + \boldsymbol{f}_i(t)\Delta t/2$;
2. $\mathcal{P}$ is evaluated using eq. 2.39 with $\boldsymbol{r}_i(t)$, $L(t)$ and $\boldsymbol{p}_i'$; then Π_V is updated: $\Pi_V(t + \Delta t/2) = \Pi_V(t) + (\mathcal{P} - P_0)\Delta t/2$;
3. $V(t + \Delta t/2) = V(t) + W^{-1}\Pi_V(t + \Delta t/2)\Delta t/2$;
4. $\boldsymbol{r}_i' = \boldsymbol{r}_i(t) + (L^2(t)/L^2(t + \Delta t/2))(\boldsymbol{p}_i'/m_i)\Delta t$;
5. $V(t + \Delta t) = V(t + \Delta t/2) + W^{-1}\Pi_V(t + \Delta t/2)\Delta t/2$, followed by two rescaling steps:
   $\boldsymbol{r}_i(t + \Delta t) = \frac{L(t+\Delta t)}{L(t)}\boldsymbol{r}_i'$,
   $\boldsymbol{p}_i'' = \frac{L(t)}{L(t+\Delta t)}\boldsymbol{p}_i'$;
6. $\mathcal{P}$ is evaluated again using 2.39 with $\boldsymbol{r}_i(t + \Delta t)$, $L(t + \Delta t)$ and $\boldsymbol{p}_i''$; then $\Pi_V(t + \Delta t) = \Pi_V(t + \Delta t/2) + (\mathcal{P} - P_0)\Delta t/2$;
7. $\boldsymbol{p}_i(t + \Delta t) = \boldsymbol{p}_i'' + \boldsymbol{f}_i(t + \Delta t)\Delta t/2$

The friction and random parts of the Langevin dynamics are represented by adding to the forces on each particle and on the volume the langevin terms:

$$\boldsymbol{f}_i(t) = -\boldsymbol{\nabla}\mathcal{U}_i(t) - \gamma \boldsymbol{p}_i(t) + \sqrt{6\gamma k_B T m}\boldsymbol{R}(t) \quad (2.40)$$

$$\Pi_V(t) = \mathcal{P}(t) - P_0 - \gamma_V \boldsymbol{\Pi}_V(t) + \sqrt{6\gamma k_B T Q}\boldsymbol{R}_V(t) \quad (2.41)$$

2.3.4 Analysis and sampling

In MD macroscopic magnitudes are obtained by averaging the microscopic variables of different configurations over time. As it has been previously stressed, the equations of motion that define the system trajectory should be ergodic, for those configurations to represent a correct sample of the working ensemble. Two time scales need to be observed in order to obtain correct averages. In MD simulations the system evolves towards states of lower free energy, and therefore, the microscopic properties of the successive configurations change with time before the steady state is reached. It is thus important to identify relaxation times, so averaging is only performed on equilibrium states. On the other hand, correlations need to be accounted for when determining the statistical error of measurements. In order to get ride of these repetitive samples, only configurations with intervals bigger than this time are recorded or used to calculate the macroscopic averages. However, since the system usually has several degrees of freedom, with different relaxation and decorrelation times, the estimation of those time scales may not be obvious. In order to reduce the poor statistics from long correlation times in specific degrees of freedom, averages are usually taken by blocks. The timeseries of a magnitude is divided in 10 blocks or so, each one containing n_i samples. The mean values from each block, $\overline{A}_i = a_i/n$, are averaged together and the standard deviations calculated in order to get a good estimate of the observable:

$$< A >= \sum_{i=1}^{n_{blocks}} \overline{A}_i / n_{blocks} \quad , \qquad \sigma(A) = \sqrt{\frac{(\overline{A}_i - < A >)^2}{n_{blocks} - 1}} \quad (2.42)$$

In general, macroscopic quantities depend on particle positions, velocities and forces, and are either calculated directly during the simulation and averaged at the end, or the configurations are recorded and the magnitudes of interest are obtained from post-simulation analysis. Temperature and pressure are typical thermodynamic quantities usually calculated in MD simulations. They do not only give important information about the system in NVE ensembles, but are crucial for thermostat and barostat dynamics in NVT and NPT. The temperature is related to the velocities of the particles via the equipartition theorem, which predicts that the kinetic energy of each degree of freedom equals $k_B T/2$:

$$< \sum_{i=1}^{N} |\boldsymbol{p}_i|^2/m_i >= 2 < \mathcal{T} >= N d k_B T \quad (2.43)$$

where N and d are the number of particles and degrees of freedom in the system, and $\mathcal{T}$ is the instantaneous kinetic energy. The pressure can also be obtained from the position, velocities and forces using the virial equation (eq. 2.39).

Diffusion describes the mobility of the particles in a solvent. In membranes, phospholipid and protein diffusion determines the contact rates and the membrane deformation, both important in signal transduction and cell communication. The diffusion constant is obtained from the mean squared displacement of the particles:

$$D = \frac{1}{6\tau} < |\boldsymbol{r}_i(t+\tau) - \boldsymbol{r}_i(t)|^2 > \tag{2.44}$$

where the average is taken over the ensamble.

In many cases, structural magnitudes are also important to characterize the system, order parameters, for example, are very useful to determine the different phases of membranes. Those magnitudes are often defined on each specific system, and need to be carefully chosen. Order parameters form a set of independent variables that can completely characterize the system. Sometimes, the whole set does not need to be defined, and only one of them can describe the features or processes of interest.

The free energy completely characterizes the thermodynamics of the system. The equilibrium states, energy barriers, and the probability of finding a specific state of the system are easily derived from it. Analytically it can only be calculated for very few and simple cases, and computationally is not trivial to obtain. When a full sample for all the degrees of freedom of a system can be performed, the free energy is obtained from the probabilities of the different states contained in the definition of the partition function, Z:

$$Z = \frac{1}{N!\Lambda^3} \int e^{-\mathcal{U}(\boldsymbol{q_0},\boldsymbol{q_1}...\boldsymbol{q_N})k_BT} \mathrm{d}\boldsymbol{q_0}\mathrm{d}\boldsymbol{q_1}...\mathrm{d}\boldsymbol{q_N} \tag{2.45}$$

where $\boldsymbol{q}_i$ are the different degrees of freedom and the coefficient before the integral corresponds to the integration of the kinetic part of the hamiltonian. The free energy is then given by:

$$F = -k_BT\ln Z \tag{2.46}$$

However, the full sampling of the system cannot usually be obtained because there are states with low probabilities that are poorly sampled or even inaccessible when energy barriers separate two regions of the configuration space (Fig. 2.4). In some cases, the system can be initialized at specific states, but in others, when those configurations are the result of the collective behavior of many particles, they can be totally inaccessible with the normal dynamics. Luckily, important information can be obtained without the need to determine fully and exactly the free energy. In many cases, the study of the system is focused along a specific reaction coordinate able to characterize the process under study. A partition function can be defined when one of the degrees of freedom is fixed:

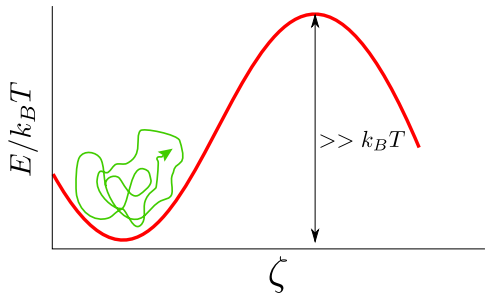

Figure 2.4: The energy barrier prevents the system to evolve to states of high energy or to cross the barrier. In this case, all the states cannot be sampled by normal dynamics.

$$Z(\boldsymbol{\zeta}) = \frac{1}{N!\Lambda^3}\int e^{-\mathcal{U}(\boldsymbol{q_0},\boldsymbol{q_1}...\boldsymbol{q_N})k_BT}\delta(\boldsymbol{\zeta}-\boldsymbol{q_0})\mathrm{d}\boldsymbol{q_0}\mathrm{d}\boldsymbol{q_1}...\mathrm{d}\boldsymbol{q_N} \tag{2.47}$$

and the equivalent to the free energy for this partition function is called the potential of mean force:

$$F(\boldsymbol{\zeta}) = -k_BT\ln Z(\boldsymbol{\zeta}) \tag{2.48}$$

When $\boldsymbol{\zeta}$ adopts all the values along the reaction coordinate, the potential of mean force can be understood as the projection of the free energy on that coordinate. Sampling can be difficult even for a single coordinate, when some states are unlikely to happen. If this is the case, a smart method to sample the system needs to be applied. **Umbrella Sampling** [121] is one of those methods, where the reaction coordinate is divided into different windows, and an artificial biasing potential $V_b(\boldsymbol{\zeta})$ is added to the system potential energy $\mathcal{U}(\mathbf{q})$ in each ot them (Fig. 2.5):

$$V' = \mathcal{U}(\mathbf{q}) + V^b(\boldsymbol{\zeta}) \tag{2.49}$$

The biasing potential shifts the original potential so that the probability to sample unaccessible states is increased. Multiple simulations are performed with different biasing umbrella potentials $V_i^b(\zeta)$ that center the sampling in different overlapping regions or windows along ζ. A reasonable choice to produce the biased ensembles is to use for every window i an harmonic function of the form:

$$V_i^b(\zeta) = \frac{1}{2}k(\zeta-\zeta_i)^2 \tag{2.50}$$

The new system with potential energy V' can be correctly sampled, but the resulting distributions are biased, $\rho_i^b(\zeta)$, and do not correspond to that of the original system, $\rho_i(\boldsymbol{\zeta})$, although, they are both related:

$$\rho_i(\boldsymbol{\zeta}) = e^{\beta V^b(\zeta)}e^{-\beta(\mathcal{U}(\boldsymbol{q})+V_i^b(\boldsymbol{\zeta}))} = e^{\beta V^b(\boldsymbol{\zeta})}\rho_i^b(\boldsymbol{\zeta}) \tag{2.51}$$

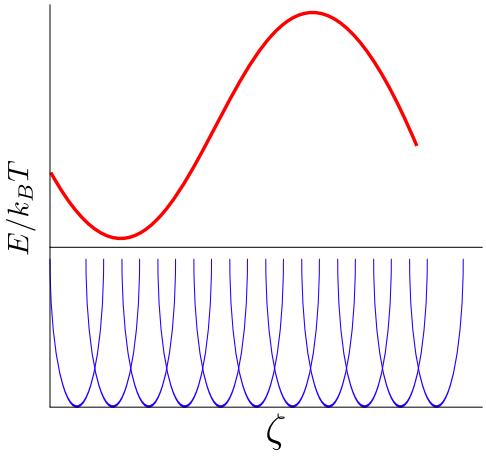

Figure 2.5: System potential along a reaction coordinate, ζ, in red, and biasing potential along the same coordinate, in blue, used together in umbrella sampling.

Since we know the biased potential, the samples can be unbiased, but to obtain a full unbiased distribution, data from all the simulations have to be combined. **The Weighted Histogram Analysis Method (WHAM)** [122] combines the samplings from the different windows and their resulting biased distributions to give an estimate of the free energy of the system. For a set of S umbrella sampling simulations where N_i samples are obtained for the i-th simulation, WHAM proceeds in the following way: these samples are discretized into M bins to determine a histogram with respect to the biasing coordinated. ρ_{ij} is the estimate of the biased probability in the j-th bin of the i-th simulation. The total biased probability distribution from the i-th simulation is then:

$$\rho_i^b(\zeta) = \sum_{j=1}^{M} \rho_{ij} \tag{2.52}$$

and the combined distribution from all the windows gives the total unbiased distribution:

$$\rho(\zeta) = \sum_{i=1}^{S} c_i(\zeta)\rho_i(\zeta) = \sum_{i=1}^{S} c_i(\zeta)e^{\beta V_i^b}\rho_i^b(\zeta) \tag{2.53}$$

where $c_i(\zeta)$ are the normalized weights, $\sum_i c_i(\zeta) = 1$. Since, in general, the total potential is not flat, the optimal weights will not be equal, and should be chosen so that the samples with higher uncertainty in their estimates of the unbiased probabilities are weighted less. A simple way to choose the weights followed by WHAM is the minimization of the variance of the estimated unbiased probability:

$$\frac{\partial \sigma^2[\rho(\zeta)]}{\partial c_i} = 0 \tag{2.54}$$

Eqs. 2.53 and 2.54 form a system of linear equations that are solved iteratively until self-consistence to obtain the weights that define the unbiased probability so the potential of

mean force along the reaction coordinate can be calculated from eq.2.48.

2.3.5 Dealing with small systems

In most cases, computer simulations are performed with a small number of molecules. Usually the total system does not need to be simulated in order to capture its main features. In liquid simulations, for example, thousands of molecules are usually enough to determine the thermodynamic properties of the system they constitute. In the case of membranes, the whole cellular membrane is not simulated to understand the phase diagram of a mixture of amphiphiles, or to determine its elastic behavior. The force calculation is the bottle neck in MD simulations, and its computing time increases linearly with the number of interactions and therefore quadratically with the number of particles. Although some techniques that will be shown in the next section can reduce this dependency, small systems decrease importantly simulation times, or even represent the only way to study complex processes. Nevertheless, a minimum size for the system is imposed by the correlation lengths or the range of the interactions. It is crucial to ensure that a small box does not introduce finite size effects on the system, that could modify its behavior. Large fluctuations close to critical points cannot take place in finite sized systems, this can inhibit phase transitions to occur or change their characteristics. Determining the optimum size of the simulation box will enhance the simulation efficiency while keeping its real features.

An important limitation of finite boxes simulations is the appearance of surface effects on the box borders. An easy way to get ride of these effects is to replicate the simulation box using **periodic boundary conditions** (PBC) [110]. The simulation box is replicated throughout the space to form an infinite lattice of replicas (Fig. 2.6). During the simulation, particles in each periodic image move symmetrically. When a particle leaves the central box, a particle from a surrounding replica enters it from the opposite border. In this way, there are no walls at the boundary of the central box and thus, no surface particles. In practice, only the central box is simulated, and the PBC are applied in two cases during the simulation: when particles cross the borders of the central box and when determining interactions with neighbors. When a particle leaves the central box on one side, PBC make it to enter back on the other. When the interactions between the particle and their neighbors are calculated, PBC use **the minimum-image convention**, so each individual particle in the simulation interacts with the closest image from its own box or from the other replicas (Fig. 2.6). This means that particles in opposite sides of the box interact via their replicas of adjacent boxes with the minimum-image convention.

2.3.6 Speeding up

The simulation time usually limits the sizes and complexity of systems that can be studied. Although since the appearance of the first commercial microprocessor in 1971, computer

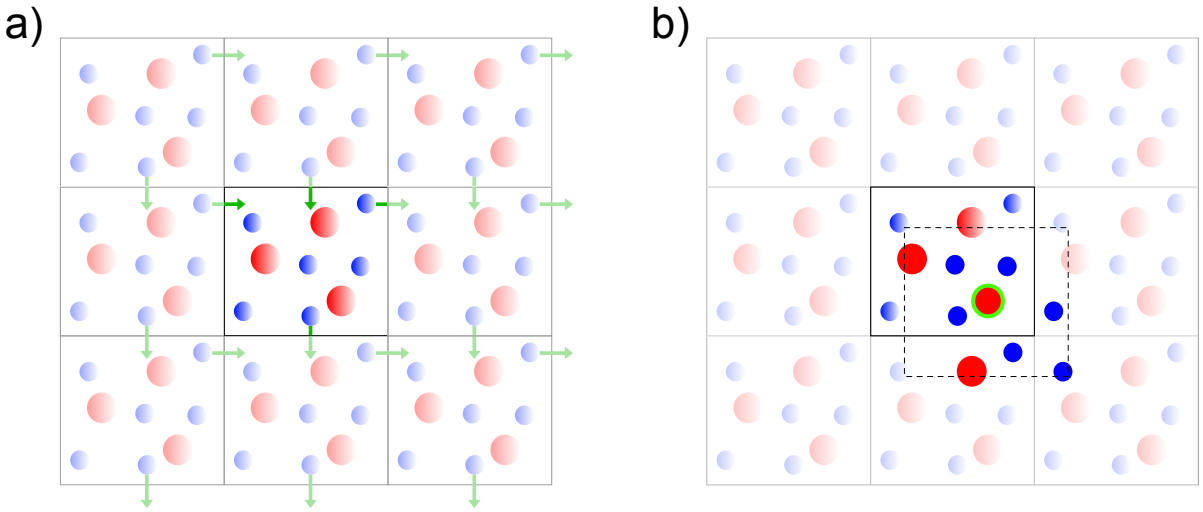

Figure 2.6: Schematic showing the replicas created around a central simulation box when applying periodic boundary conditions in two dimensions. The particles in the central box are shadowed while those in the replicas are transparent. a) Particles movement inside and outside the central box. When a particle exits the central box through one border, a symmetric particle from an opposite replica enters the central box from the other side. b) Schematic showing the minimum-image convention, the interactions are calculated between the tagged particle and those in solid color, which are the nearest representatives of system particles in the closest replica.

capacity has expanded rapidly, and today big supercomputer centers are available in many research centers, an efficient algorithm is still crucial to make the most of the actual technology and make possible to increase system size and detail.

Apart from determining an optimal size for the simulation box, there are other techniques that can be used to speed up simulations. As it was mentioned before, the bottleneck in MD simulations is the force calculation, since it is a function of N^2, while position and velocity update depends only on N. The first approximation done to calculate the forces on a given particle is to set a maximum interacting radius for a specific potential, beyond which particles are supposed to be non-interacting. The largest contribution to the potential and forces comes from neighbors close to the molecule of interest, applying a cutoff, r_c, means setting the pair potential $v(r)$ to zero for $r >= r_c$. The introduction of this cutoff should be a small perturbation on the thermodynamic properties. Since most potentials approach to zero asymptotically, it is not evident where to set the cutoff. Usually it is set at a distance where the potential, the force, and its derivative are close enough to zero so the accuracy of the potential is not affected. The cutoff distance must also be no greater than half the box size for consistency with the minimum image convention. In short-ranged potentials, this cutoff can be set directly, and the value of the potential energy at the cutoff is substracted from the total energy, while for long-range potentials, as Coulomb interactions, some approximations

need to be used [111].

Although cutoffs allow to neglect interactions at long ranges, the distance between particles need to be computed every step to identify the interacting particles, this still implies a calculation of the order of N^2 at every step. One common technique to reduce this calculation time is the use of **neighbor lists**. A sphere of radius r_n slightly greater than the maximum potential cutoff, r_c, is set on every particle in the simulation box, $r_n = \max(r_c^i) + \delta$. The neighbors of a particle, those that lie inside its surrounding sphere, are recorded (Fig. 2.7). For a small number of steps n_n, the interactions are only calculated among the neighbors registered on the list. After n_n steps the list is updated again with the actual positions. The values of n_n and δ are related, the bigger δ is, the more time the neighbor list can be used without needing to update it, because no new interacting pair of particles has moved close. The best value of those parameters is a balance between a reduced neighbor list that allows to calculate fewer interactions, and a seldom building of that list, which is also time consuming.

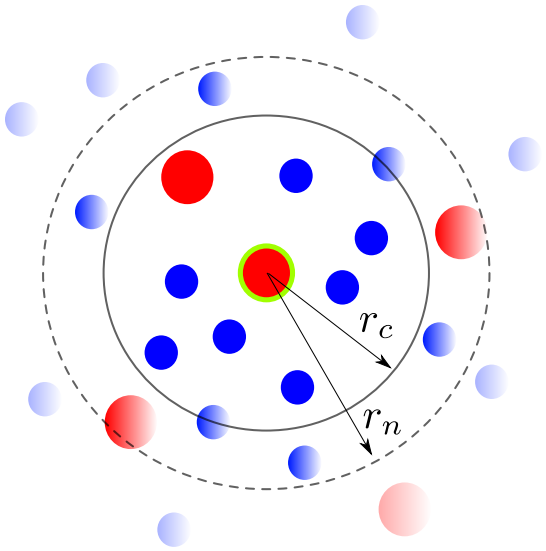

Figure 2.7: Schematic showing the neighbor list construction in two dimensions. A circle of radius r_n is set around the tagged particle. All the particles lying inside that circle are included in the neighbor list of the tagged particle, although only the particles at a distance lower than r_c will contribute with an interaction different from zero.

The neighbor list construction can be accelerated with the use of **linked cells**. Instead of calculating the distance with all the particles in the simulation box to determine the neighbors, the simulation box is divided into a regular lattice of MxMxM cells , and only the distances between particles in adjacent cells is checked (Fig. 2.8). The cells are chosen so that the size of the cell $l = L/M$ is greater that the cutoff distance of the neighbor list, r_n. In systems with $L >> l$ this is a great improvement in the neighbor list calculation speed, since only particles at distances smaller than $2l$ are checked to build the list. In this way,

the number of computed interactions between particles is highly reduced since each particle does not interact with the rest of the system but only with its neighbors. Therefore, the force calculation using neighbors list and linked cells grows linearly with the number of particles, instead of quadratically.

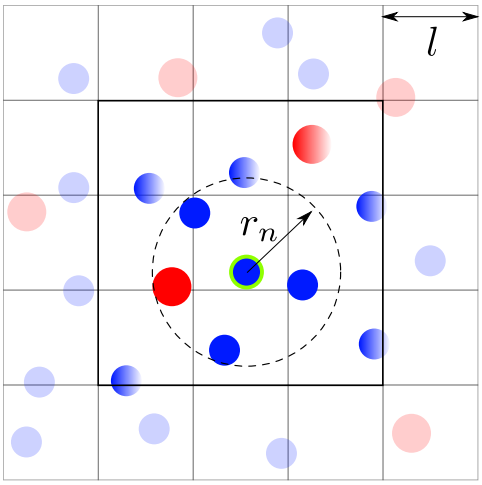

Figure 2.8: Schematic showing the linked cells construction in two dimensions. The system is divided into a regular lattice of boxes of side l, the particles that will be added to the neighbor list of a tagged particle are selected only from the adjacent boxes.

Even with the development of highly optimized algorithms, molecular simulations have always been limited in size and timescale by the available computing resources. Due to the highly parallelizable algorithms typical of these simulations, where many calculations can be carried out simultaneously, MD and MC have benefited from the appearance in the 1990s of supercomputers with thousands of processors. MD simulations are in fact inherently parallel: for a given configuration, the force calculation, and the position and velocity updates can be computed independently, and therefore simultaneously for every particle. In principle, the time consumed in the parallel parts of the algorithm should scale with the number of processors, so the use of 100 processors would allow to run the same simulation 100 times longer. In practice, this is not the case, and there is an optimum usage of resources. The scaling factor is reduced due to, among others, communications delays between processors and synchronization problems. However, parallel codes are the norm today, and they have boosted the limits of standard molecular simulations.

Parallelization, however, is not an easy task, and usually complex approaches, sometimes related to the specific architecture of a supercomputer, are followed. The need of very specific and technical coding resulted in the development of simulation packages with optimized algorithms that are widely used today, such as GROMACS [123],NAMD [124], CHARMM [125] and several others. While these packages have become a very useful tool for atomistic MC and MD simulations, they are usually not so efficient for the more flexible coarse grained

interactions.

Recently, **graphic processing units (GPUs) computing** has been introduced to the research community with an enormous success. In the last decades, the capacity of the GPUs grew faster than that of the CPUs. Propelled by the industry of entertainment, they needed to develop a tremendous computational power, required for the interactive rendering of complex three dimensional images and visual simulations typical of video games. Until recently, GPUs were used solely for their ability to process images and geometric information at blistering speeds. The parallel nature of many computer graphics algorithms naturally led the design of GPUs toward hardware and software architectures that could exploit this parallelism to achieve high performance [126]. This has resulted in GPU designs that contain from hundreds to thousands of processing subunits, fed by multiple independent parallel memory systems. Molecular simulations were not able to exploit this parallel architecture and high performance until Nvidia developed the Compute Unified Device Architecture (CUDA) [127], a computing engine for Nvidia GPUs that is accessible to software developers. Programmers use *c* language with some Nvidia extensions that allow to access to the different memories and parallelization capabilities of GPUs.

Current GPU hardware designs still have limitations. While the large number of parallel processing units and memory systems accounts for the tremendous computational power of GPUs, it is also responsible for constraining the practical size of memories, caches, and register sets available to each processor. [126]. Nevertheless, this is still a fast evolving field both in hardware and software design. Programs and codes are still highly dependent of architecture features, but GPUs are spreading very fast because they allow to obtain the same parallelization efficiency in a nonexpensive personal computer than in massive supercomputers with multiple processors. New packages, such as HOOMD [128, 129], have also been developed for the simulation of MD in GPUs, and older packages designed to be used in CPUs, such as GROMACS or NAMD, are now upgraded to be able to use them for GPU computing.

2.4 Continuum models for membranes

Particle-based models are widely used in the simulation of biological systems, coarse graining is specially useful in the study of membranes due to their mesoscopic nature, where collective effect of the molecules determine the macroscopic features of the bilayer. However, elastic properties can also be described by continuum theories. Since the thickness of the lipid bilayer is much smaller than the scale of the whole bilayer, membranes can be seen macroscopically as two dimensional elastic surfaces subjected to a differential geometry description. Their curvature energy is described by the Helfrich Hamiltonian [130] :

$$H = \int \mathrm{d}A \left(\sigma + \frac{\kappa}{2}(2C - C_0)^2 + \kappa_G K\right) \tag{2.55}$$

where σ is the surface tension, κ and κ_G are the bending rigidity and the Gaussian curvature modulus respectively, C and K are the mean and Gaussian curvatures, and C_0 is the spontaneous curvature. The second term in the integrand measures the quadratic deviation of the total curvature from the spontaneous curvature C_0. For a symmetric and tensionless membrane, C_0 and σ vanish. Assuming that the Gaussian curvature modulus κ_G is invariant throughout the membrane, the last term in 2.55 is constant under the Gauss-Bonnet theorem [131], so as long as neither boundary nor topology changes occur, the Gaussian term is rigorously irrelevant. Therefore, the Helfrich Hamiltonian for the description of biomembranes can be reduced to:

$$H = \int \mathrm{d}A \left(\sigma + 2\kappa C^2\right) = \int \mathrm{d}A \left(\sigma + \frac{\kappa}{2}(\kappa_1 + \kappa_2)^2\right) \tag{2.56}$$

where κ_1 and κ_2 are the principal curvatures. The important coefficients that describe the membrane are thus, the bending rigidity, that describes the resistance of the system to bending, and the surface tension σ, which takes into account the contribution of the bilayer total area A, to the system energy.

Using the Monge representation, the shape of the bilayer can be described by the height function $h(x, y)$ which measures the distance of its midsurface from the reference x-y plane. With this representation, the area element and the mean curvature take simple forms [13]:

$$\mathrm{d}A = \mathrm{d}x\mathrm{d}y\sqrt{1 + (\boldsymbol{\nabla} h(x,y))^2} \tag{2.57}$$

$$C = \frac{1}{2}\left(\frac{\partial^2 h(x,y)}{\partial x^2} + \frac{\partial^2 h(x,y)}{\partial y^2}\right) \tag{2.58}$$

and the classical elasticity Hamiltonian becomes:

$$\begin{aligned} H &= \int_A \mathrm{d}x\mathrm{d}y \left[\frac{\sigma}{2}(\boldsymbol{\nabla} h(x,y))^2 + \frac{\kappa}{2}(\boldsymbol{\nabla}^2 h(x,y))^2\right] \\ &= \frac{1}{2L^2}\sum_{\boldsymbol{q}} \left[\sigma q^2 + \kappa q^4\right] |\tilde{h}(\boldsymbol{q})|^2 \end{aligned} \tag{2.59}$$

where $\tilde{h}(\boldsymbol{q})$ are the Fourier coefficients of $h(x, y)$ and L is the side dimension of the projected area for a square membrane:

$$\tilde{h}(\boldsymbol{q}) = \frac{1}{L^2}\int_A h(x,y) e^{i\boldsymbol{r}\cdot\boldsymbol{q}} dxdy \tag{2.60}$$

A quantitative measure for the shape fluctuations of the bilayer membrane is given by the fluctuation spectrum $S(\boldsymbol{q}) = |\tilde{h}(\boldsymbol{q})|^2$, that together with the equipartition theorem, takes the form:

$$S(\boldsymbol{q}) = \frac{k_B T}{L^2(\sigma \boldsymbol{q}^2 + \kappa \boldsymbol{q}^4)} \tag{2.61}$$

The fluctuation spectrum has been obtained under the hypothesis that the bilayer can be approximated to a two dimensional surface, so its width has been neglected. Therefore eq. 2.61 holds only for wave numbers smaller than $q_{max} \simeq 2\pi/d$ where d is the bilayer thickness. Simulation results can be fitted to the fluctuation spectrum to obtain the bending modulus, κ, and tension σ.

Since for wave vectors smaller than $q \simeq \sqrt{\sigma/\kappa}$ the fluctuations are tension dominated, it is best to simulate at zero tension in order to avoid unnecessary damping of the most relevant modes [132]. The lateral size of the simulation box can be adjusted in order to obtain a bilayer with vanishing tension [133] , and data are fitted to eq. 2.61 with $\sigma = 0$. There are some limitations and difficulties when calculating κ using thermal fluctuations. First of all, the minimum wave number is determined by the size of the box, so in order to measure large fluctuations, large membranes with their associated computational cost, need to be simulated. Bending moduli in stiff membranes are also difficult to obtain. From eq. 2.61, it is evident that large values of κ lead to very small amplitudes that decay with increasing wave vector, so large statistics are needed to resolve the spectrum. Some techniques have been presented to calculate the bending coefficient by measuring the response of the membrane when imposing large curvatures [132, 134]

When measuring the fluctuation spectrum of simulated membranes, the surface of the membrane needs to be well defined. From configurations obtained in simulations, only the positions of the molecules forming the membrane are obtained, but not a geometrical function. A surface is usually defined from the positions using Delaunay triangulation or Voronoi tessellations [135] in order to perform the Fourier analysis of the fluctuations.

3 Models

3.1 Modeling virus budding

Virus budding involves processes from the microscopic to the macroscopic scale. The description of the virus subunits and membrane components at the molecular level is necessary to understand the binding between virus and membrane, or between the capsid subunits during assembly. In addition, the membrane lipid composition and the atomistic description of the molecules determine the macroscopic properties, such as the bending coefficient or the area per lipid.

On the other hand, virus budding implies membrane deformations in a scale of hundreds of nm to μm generated by the collective behavior of hundreds of thousands of lipids. The membrane response at this scale has been widely treated with continuum elastic models.

In order to connect these two scales, models in the mesoscopic range that allow to identify the microscopic characteristics responsible of the macroscopic behavior are used. Since these models do not describe the system at atomistic detail, the microscopic interactions are usually represented via effective interactions, that reproduce the main features of the system behavior.

In this context of complex systems involving different scales, minimal models are particularly useful in addressing the macroscopic system behavior, while connecting it to the microscopic properties via a minimum set of variables, that are enough to characterize the features under study. Coarse grained models are ideal candidates to describe such approximations but require a deep understanding of the system in order to define an adequate set of interactions capable of reproducing the interesting processes while shedding light on the origin and nature of them.

3.2 Membrane Models

As it was stressed in Chapter 1, biological membranes are multicomponent systems, constituted by a number of different components, that show a highly complex behavior related to their many biological functions. However, when designing a simulation, or a membrane model, only the important properties of the system under study are usually represented. This simplification allows to focus on the desired properties of the system and to easily identify the effects of the reduced number of elements.

As we commented above, the high efficiency of these reduced models also allows to study otherwise inaccessible time and length scales. This is why instead of reproducing complex real membranes, coarse grained models are usually based on simple model bilayers, where individual elements or properties are added in order to study a specific process.

3.2.1 Explicit solvent models

Coarse grained models with explicit representation of the solvent have been very successful in the simulation of stable membranes, and despite their simplicity they have been remarkable accurate at reproducing the structural, thermodynamic, and mechanical properties of phospholipid aggregates [109].

In these models, the solvent is usually represent by a fluid of interacting beads. Goetz and Lipowsky [136] presented a simple model based on a Lennard Jones fluid of solvent beads and amphiphilic molecules. The amphiphiles were made by one or two hydrophobic tails attached to a hydrophilic head. The solvent beads and the amphiphilic head were represented by hydrophilic beads while the tails of the surfactants were modeled by hydrophobic beads (Fig. 3.1). The beads constituting an amphiphilic molecule are linked together by harmonic springs and a bending potential that controls the geometry and stiffness of the tails. The interactions between one hydrophobic and one hydrophilic molecule are modeled by a repulsive soft core potential $(\sim 1/r^9)$, while all other interactions are modeled by Lennard Jones potentials. Depending on the lipid concentration, spontaneous assembly into micelles or fluid bilayer was observed with this model. Varying the length and geometry of the lipid tails, the effect of the amphiphile architecture on the structure and elastic properties of the membranes were observed [136, 137].

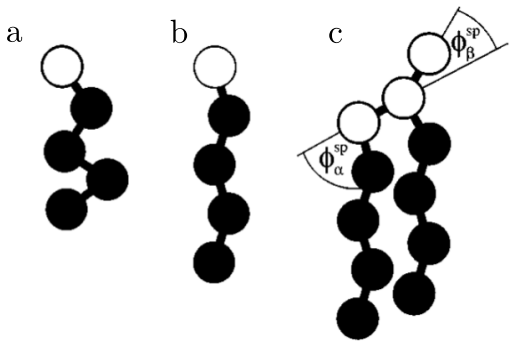

Figure 3.1: Three types of surfactants represented according to the model proposed by Goetz and Lipowsky. Hydrophobic beads are represented in white and hydrophobic ones in black. Amphiphiles with a) flexible tail, b) semi-flexible tail and c) double tail. Figure adapted from [136]

A similar but more simple model where only two type of beads are used was developed

by Stevens [138]. The solvent particles are represented by single hydrophilic beads of the same type of those for the head groups. This model was successfully applied to the study of thermotropic phase behavior of bilayers formed by lipids of varying length [138], and domain formation in mixed bilayers of long and short amphiphiles [139].

A more sophisticated model was introduced by Marrink [140] with the aim of representing the chemical nature of the molecular components . A set of bead types was used to build molecules with different architecture. Each bead has the volume of four water molecules on average, and only four interaction types are represented: polar, nonpolar, apolar and charged (Fig. 3.2). Polar sites represent neutral groups of atoms that would easily dissolve into water, apolar sites represent hydrophobic moieties, and nonpolar groups are used for mixed groups which are partly polar, while charged sites are reserved for ionized groups. Different subtypes are considered within a type of interaction site depending on the chemical nature of the atoms comprising the group, this allows for fine tuning of a specific molecule. An alkane/water system was used as a reference system to calibrate the length and energy scales of the Lennard Jones parameters, while those for the bonded forces were derived by fitting against atomistic simulations. This model reproduces the self-assembly of phospholipids into a bilayer, and the density profiles along the membrane normal are in good agreement with results from atomistic simulations [140]. This model was also used to reproduce and study the gel phase [141], to investigate vesicle structure and dynamics [142], and to study the transition from a multilamellar to a nonlamellar inverted hexagonal phase [143]

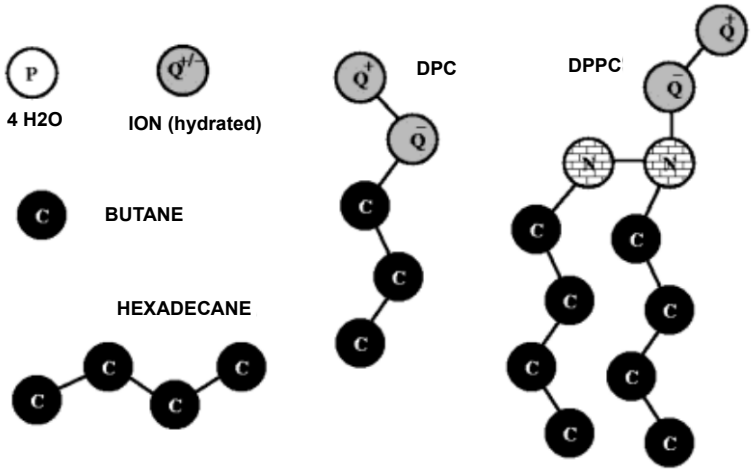

Figure 3.2: Representation in the model proposed by Marrink for water, ions, butane, hexadecane, DPC (dodecylphosphocholine), and DPPC (dipalmi- toylphosphatidylcholine). All atoms except nearest neighbors interact through a Lennard-Jones potential. Four main atom types are distinguished: polar (P), nonpolar (N), apolar (C), and charged (Q). Figure adapted from [140]

Many other explicit solvent coarse grained models have been developed and used to ad-

dress different features of amphiphile systems, such as the differences in the phase behavior between single-tail [100] and double tail lipids [99, 138], the effect of alcohol [144] and cholesterol [145] on lipid bilayers, the interaction between membrane proteins and bilayers [104, 146], or vesicle formation and fusion [142, 147] among others.

3.2.2 Implicit solvent models

The explicit representation of water allows the definition of the correct interaction between polar and nonpolar moieties of phospholipids and water that result in the hydrophobic effect. Those models represent also a more realistic and accurate description of amphiphilic systems. However, the inclusion of the solvent accounts for an important fraction of the simulated molecules and therefore, most of the simulation time is spent on it. Implicit solvent models improve the simulation efficiency by removing the solvent, while preserving the overall physical behavior. In these models the water is usually replaced by effective forces acting on the lipids in order to form stable aggregates.

One of the first coarse grained implicit solvent model was proposed by Sintes and Baumgärtner [148]. The heads of the lipid molecules were tethered to a continuous sheet where the molecules could diffuse freely. With this model the values of average mean area per lipid and the average tilt angle were in agreement with experimental values. An important limitation of this approach is that the sheets that confine the amphiphile heads do not represent the deformations and fluctuations typical of lipid bilayers. To overcome this constraint, Laradji [149] developed a model where the sheet fluctuated according to the Helfrich elastic model for MC simulations. General structural properties and equilibrium states can be studied with this model, but the dynamics of the system cannot be reproduced, since the thermal fluctuations do not arise from the collective response of the phospholipids as a consequence of the nature of the interactions between lipids.

The most natural and easiest way to implement the hydrophobic effect is through effective interactions between lipids, however, many developers found an important drawback of this approach; the use of standard attractive potentials between lipids did not yield fluid but crystallize membranes. Many attempts were needed before Drouffe et al [150] and Noguchi and Takasu [107] found that the use of multi body potentials were necessary to reproduce the fluid regime. Drouffe's model consisted on a bilayer made of anisotropic spheres interacting with an orientation dependent pairwise potential and an isotropic many body potential (Fig. 3.3a). This model is extremely efficient and was used to self-assemble a vesicle ten years before a much smaller atomistic bilayer was assembled [151]. Noguchi's model changed the anisotropic sphere representation of a lipid for a a rigid chain of three beads, one for the head and two for the tail (Fig. 3.3b). The effective many body potential was applied between the hydrophobic beads. This model was used to study the self assembly of vesicles, vesicle fusion [107], and the adhesion of nanoparticles to vesicles [152] among others.

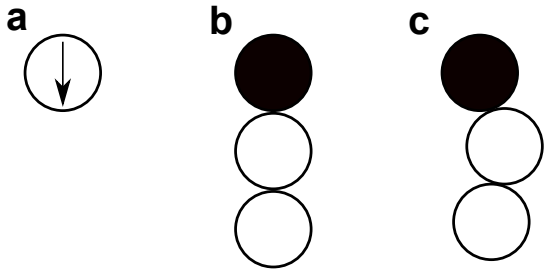

Figure 3.3: Amphiphile representation for different implicit solvent models. a) Lipids represented by oriented spheres in the model proposed by Drouffe et al. b) Rigid chains of three bears used in bot Noguchi and Takasu model and Farago pairwise model. c) Three-beads amphiphile from Cooke model with intramolecular soft interactions.

Farago was the first to show that many body potentials were not essential to reproduce fluid membranes [153]. From a practical stand point, pair potentials allow for more efficient code and simplified algorithms, and standard molecular simulation packages only support pairwise interactions. In the Farago model each lipid is represented as a rigid chain of three spherical particles, one for the head and two for the tail as in Noguchi model. The beads on different molecules interact via LJ-type potentials; with purely repulsive interactions between hydrophilic and hydrophobic particles. The coefficients for these interactions are different for each combination of beads, so 6 sets of coefficients need to be specified in the force field. This model was used to study pores and elastic properties of membranes [153, 154]. The values of the bending rigidity obtained for the membranes, $\sim 50 k_B T$, were better than those from the preceding models but still high compared to the experimental values.

A much simpler model was proposed by Cooke, Kramer and Deserno [93] with just two types of interactions, a repulsive one to avoid overlapping between beads, and an attractive one between tails to yield aggregates (Fig. 3.3c). They also showed that fluidity was a function of the range of the attractive interaction between tails. Since a liquid phase needs both sufficient cohesion in order to compete with the much larger entropy of the gas phase, and to retain sufficient entropy to compete with the solid, short range attractive potentials can only stabilize a condensed bilayer in the gel phase [91]. The bilayers obtained with this model self assemble spontaneously (Fig. 3.4) and have bending rigidities between 3 and 30 $k_B T$, similar to those observed in biological membranes.

Many other coarse grained models of biological membranes have been developed [109, 151], however the focus will be set on presenting the Cooke Model in full detail since it will be extensively used later in this work.

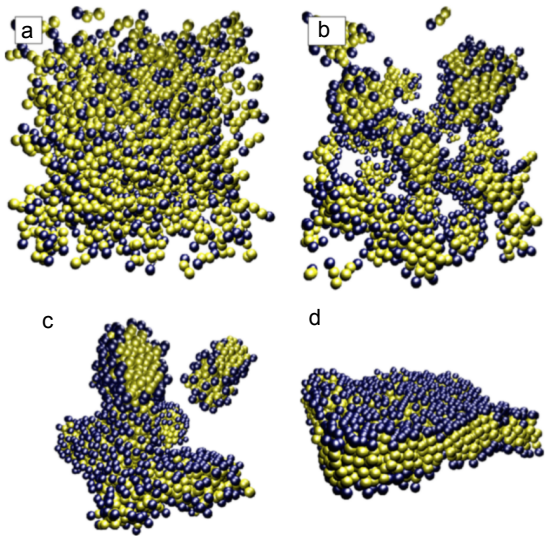

Figure 3.4: Self-assembly sequence of a system with 1000 lipids into a bilayer using the Cooke Model. A random gas of lipids (a), forms small clusters (b), that aggregate together (c), and eventually form a box-spanning bilayer (d). Figure adapted from [93]

Cooke Model

In this model, the amphiphilic lipids comprising the membrane are represented by one head bead and two tail beads that interact via repulsive WCA potentials [155], 3.1

$$V_{\text{rep}}(r) = \begin{cases} 4\epsilon_0 \left[\left(\frac{b}{r}\right)^{12} - \left(\frac{b}{r}\right)^{6} + \frac{1}{4} \right] & ; r \leq r_c \\ 0 & ; r > r_c \end{cases} \tag{3.1}$$

with $r_c = 2^{1/6} b$ and b is chosen to ensure an effective cylindrical lipid shape: $b_{\text{head-head}} = b_{\text{head-tail}} = 0.95\sigma$ and $b_{\text{tail-tail}} = \sigma$, where σ will turn out to be the typical distance between beads within a model lipid molecule. A cylindrical lipid shape is supported best when the head bead is chosen a tiny bit smaller than the tail beads. The reason is that the cohesion, which will later be used to drive aggregation, pulls the tails together but not the heads [91]

The beads belonging to a given lipid are connected through FENE bonds, 3.2 [156], and the linearity of the molecule is achieved via a harmonic spring with rest length 4σ between the first and the third bead, 3.3

$$V_{\text{bond}}(r) = -\frac{1}{2}\kappa_{\text{bond}} r_\infty^2 \ln\left[1 - (r/r_\infty)^2\right] \tag{3.2}$$

where $r_\infty = 1.5\sigma$

$$V_{\text{bend}}(r) = \frac{1}{2}\kappa_{\text{bend}} (r - 4\sigma)^2 \tag{3.3}$$

Since this is an implicit solvent model, the hydrophobicity is represented by an attractive interaction, 3.4, between all tail beads.

$$V_{\text{attr}}(r) = \begin{cases} -\epsilon_0 & ; \quad r < r_c \\ -\epsilon_0 \cos^2 \frac{\pi(r-r_c)}{2\omega_c} & ; \quad r_c \leq r \leq r_c + \omega_c \\ 0 & ; \quad r > r_c + \omega_c \end{cases} \tag{3.4}$$

These interactions lead to the spontaneous self assembly of the lipids into a bilayer (Fig. 3.4). Depending on the temperature and the range of the hydrophobic potential, three different phases are observed (Fig. 3.5). At low temperatures, the bilayer forms a gel state showing a high degree of order, while for higher temperatures, an increase in entropy results in fluid membranes for a broad range of the parameters. Even higher temperatures give place to unstable bilayers. Physical properties of the membrane such as fluidity, area per molecule and bending rigidity are therefore easily tuned via ω_c (Fig. 3.6).

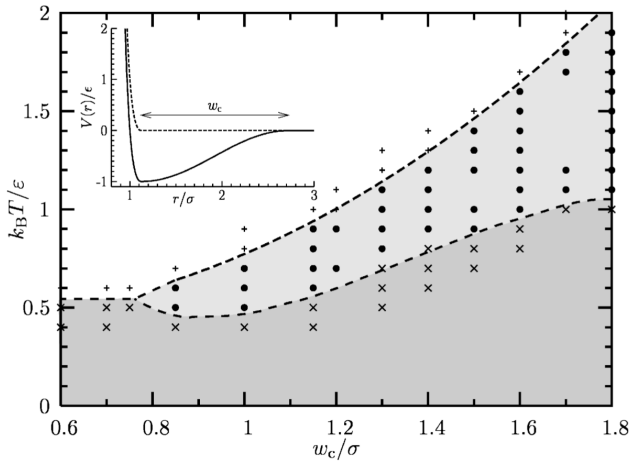

Figure 3.5: Phase diagram for the Cooke model at zero lateral tension as a function of the potential width ω_c and temperature. Each symbol corresponds to one simulation and identifies different bilayer phases: (×) gel, (•) fluid, and (+) unstable. The lines are merely guides to the eye. The inset shows the pair potential between tail lipids (solid line) and the purely repulsive head-head and head-tail interactions (dashed line). Figure adapted from [93]

Among the benefits of this model are its simplicity, its robustness towards membrane formation, the broad range of fluidity and the tunning capability of the membrane physical properties. Moreover, diffusivity within the membrane, density, and bending rigidity are in good agreement with values of these parameters measured for biological membranes.

The Cooke model has been applied to the study of membrane curvature as a function of lipid phase [157], the aggregation and vesiculation of membrane proteins by curvature mediated interactions [158], or the interactions of lipid bilayers with supports [159].

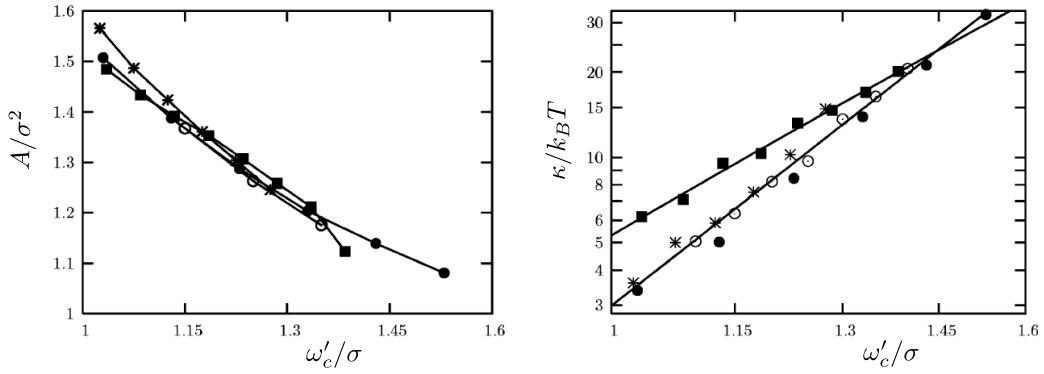

Figure 3.6: Bilayer properties. Area per lipid (a), and bending energy (b) as a function of the rescaled attractive potential $\omega_c' = 1 + \omega_c - \omega_c^{lg}$ for four different isotherms: $k_BT = 0.6$ (filled squares), $k_BT = 0.8$ (asterisks), $k_BT = 1.0$ (empty circles) and $k_BT = 1.1$ (filled circles). ω_c^{lg} is the value of ω_c on the liquid-gas transition line. Figure adapted from [93]

3.3 Virus Models

Virus assembly represents a beautiful example of how collective interactions drive the formation of complex geometrical structures, this is enough to attract many researchers to develop models to describe this phenomenon. However, virus modeling goes beyond the study of elegant toy models that address general mathematical or physical concepts, and provides useful information to characterize the actual biological process.

Experimentally, it has been difficult to characterize the assembly pathways because of the low levels of intermediate states found in capsid assembly. Detailed equilibrium structures have also been unaccessible for long times due to the small scales, of the order of nm. Although the development of cryo-electron microscopy and fluorescent labeling has improved the experimental study of virus assembly, theoretical and computational models have been very helpful at elucidating important features of assembly pathways and mechanisms.

Although in general virus assembly involves the coordinate action of capsid proteins, nucleic acid encapsulation, acquisition of membrane coats, and maturation of the virions, with the possible mediation by other viral proteins or cell machinery, it has been observed that capsids can form spontaneously in vitro [160]. So many models have focused on just representing the assembly of the capsomers [161].

Protein-protein interactions between capsomers that drive assembly are a combination of hydrophobic, electrostatic, van der Waals and hydrogen bonding. All these interactions are short-ranged under assembly conditions. Van der Waals interactions and hydrogen bonds operate on length scales of a few Å, electrostatic interactions are screened on the scale of the Debye length, of the order of nm, and hydrophobic interactions lie on the scale of 0.1-1nm [161]. Since capsomer association is driven by hydrophobic interactions, solvent plays

an important role in virus assembly. However, implicit solvent models have been successful at incorporating the solvent effects in effective interaction potentials between the subunits. Other important effects of the solvent, such as the random buffeting it provides that helps annealing, can be incorporated via Brownian dynamics. Finally, regarding the hydrodynamic coupling, there are no evident differences in the assembly behavior between the explicit solvent and Brownian dynamics simulations [161].

Due to the large ranges of length and time scales (Å–μm, ps–minute), and the fact that capsid proteins typically have several hundred of amino acids, assembly is usually described with coarse grained models. However, Freddolino et al. [162] have performed an all-atom simulation of the assembled STMV virus, and recently, systematically coarse-graining from atomistic simulations have been applied to estimate subunit positions and orientations from cryo-electron microscopy images of the immature HIV capsids [163]

Coarse grained models usually represent the capsid subunits highly simplified with an excluded volume geometry and orientation-dependent attractions designed such that the lowest energy structures are shells with icosahedral symmetry. Among coarse grained models different approaches have been followed. Some of the most popular are the patchy-sphere models, where subunits have spherically symmetric excluded volume and patches with short ranged and directional specific attractions arranged such that the lowest energy configuration corresponds to a particular target structure [164–166] (Fig. 3.7a). The first dynamical simulation of virus assembly was performed with DPD by Schwartz and co-workers [165] with a patchy-sphere model representing a 60 subunits T=1 closed shell. Hagan and Chandler [164] also used a model of this kind, to obtain the first statistical description of assembly into icosahedral shells. They obtained a kinetic phase diagram with the dominant assembly products as a function of subunit concentration, subunit-subunit interaction strength, and the orientational specificity of subunit interactions.

Other types of models consider a subunit comprised of spherically symmetric beads arranged to have short ranged attractions and excluded volume geometries that reproduce protein geometries seen in capsid structures (Fig. 3.7b), such as trapezoidal shape [167, 169, 170] which is consistent with that of capsid proteins with the beta-barrel architecture or triangular subunits that form icosahedral shells [171–175]. Rapaport used models of this kind with triangular [174] and trapezoidal [167] subunits to suggest that a simple interaction potential could direct assembly of well-formed capsids. Elrad et al. investigated the encapsulation of a polymer by an icosahedral virus with a model of triangular subunits [172].

Some models have also represented the capsid with triangulated networks [168, 176], with an elastic energy governing the bond lengths and angles (Fig. 3.7c). The subunits are not set explicitly in solution, but an irreversibly growth model is used, where units are added to a capsid, or discrete changes in the network geommetry are applied according to the probability obtained from the system Hamiltonian.

Another approach that offers interesting results is the simulation of subunits confined to a

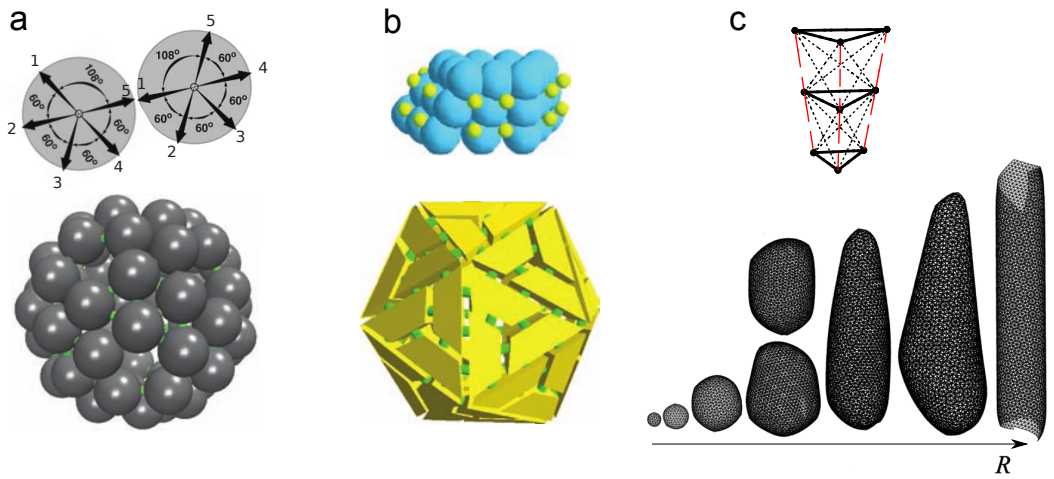

Figure 3.7: Subunit configurations for different capsid models. (a) A patchy-sphere model with the pentavalent subunit interaction geometry of a T =1 capsid, but spherically symmetric excluded volume [164]. In the top image, two interacting subunits are shown, with numbered arrows indicating the locations of the 5 distinct attractive patches. The lower image shows an assembled capsid. (b) An extended subunit representation of a T =1 capsid [167]. In the top image, the large cyan spheres experience repulsive excluded volume interactions while small yellow spheres experience attractive interactions. The lower image shows a complete capsid, with subunits reduced in size for visibility. (c) Triangulated network model. In the top image, a subunit in the form of a triangular prism is shown [168]. The solid and dashed lines represent springs with different spring constant. The lower image shows structures obtained with the triangular prism subunit as a function of the spontaneous radius of curvature, R. Figure adapted from [161, 168]

surface. Zandi et al. [177, 178] studied shells comprised of two species of discs, representing pentamers and hexamers confined to a spherical surface, and found that the Caspar and Klug geometries correspond to minimum free energy configurations for appropriate size ratios of the discs. Luque, Zandi and Reguera [179] in turn minimized the energy of those types of disks confined to a prolate surface and obtained structures consistent with prolate or bacilliform capsids.

Despite the diversity of these models, general characteristics of virus assembly studied experimentally and theoretically are also observed in most of these models. In general the phase behavior of assembly can be separated into five regimes [161]:

- No assembly at equilibrium. The interactions driving assembly are too weak to overcome the rotational and mixing entropy of free subunits, and all subunits are present as free monomers or dimers in equilibrium. This regime corresponds to subunit concentrations below a critical value $\rho < \rho^*$.

- No assembly on relevant time scales due to a nucleation barrier. For concentrations sufficiently close to the critical value, $\rho \gtrsim \rho^*$, nucleation barriers are prohibitive. Thus assembly is not seen on timescales that are accessible to simulations at these concentrations

- Productive assembly. For moderate parameter values initial nucleation barriers are small enough such that finite-time assembly yields can be quite large, with 90% of subunits assembled.

- Free subunit starvation kinetic trap. For larger subunit concentrations, new nuclei form faster than existing ones complete assembly, and free subunits are depleted before most capsids finish assembling. The system then becomes kinetically trapped.

- Malformed capsids. Subunits form bonds which do not belong to the equilibrium state geometry, an new subunits get trapped into malformed growing capsids. In some cases the capsids are able to close to make shells which lack icosahedral symmetry, whereas in other cases the curvature is disrupted significantly enough that they form long open structures.

An interesting coarse grained model for capsid assembly that lies somehow out of the previous classifications was proposed by Wales [180]. The very simple subunits are pyramids of pentagonal or hexagonal base, represented by rigid bodies defined just by five of six sites. The high simplicity of this model, together with its robustness towards assembly confer this model with ideal features to modelize the assembly and budding of virus particles on a membrane.

3.3.1 Wales Model

Here, an updated version of the first Wales model is described [181]. In this case, the subunits are represented by rigid, multi-site hexagonal and pentagonal pyramids. Depending on its pentameric or hexameric symmetry, each pyramid has five or six sites arranged as a regular polygon of radius r_b in the basal plane, and two apex sites on opposite sides of the plane at distance h. The attractive interaction between capsomers is introduced via a Morse potential between basal sites of different rigid bodies. The inherent curvature of the capsid and excluded volume of the capsomers are determined by the repulsive interaction between the upper apex sites. Finally, the lower apex sites also interact repulsively with the upper apex sites of other capsomers to ensure correct arrangement with the right subunits orientation. The total interaction energy between two pyramids is then:

$$U = \epsilon_{rep}\left(\frac{\sigma}{R_{ax}}\right)^{12} + \epsilon_{rep}\sum_{i=1}^{2}\left(\frac{\sigma}{R_i}\right)^{12} + \epsilon_{att}\sum_j\sum_k\left(e^{\rho(1-R_{jk}/R_e)} - 2\right)e^{\rho(1-R_{jk}/Re)} \tag{3.5}$$

where R_{ax} is the distance between the two primary apex sites, R_i is the distance between a secondary apex site and the primary apex site on the other capsomer, and R_{jk} is the distance between j and k, j running over the basal sites on the first capsomer , while k runs over the basal sites on the second. R_e is the equilibrium pair distance for the Morse interaction. The Lennard Jones range parameter σ is defined as

$$\sigma = R_e + r\sqrt{\frac{\sqrt{5}+5}{2}} \tag{3.6}$$

where r is the radius at the base of the capsomer (Fig. 3.8). This range parameter was chosen to minimize repulsion between two adjacent pentameric capsomers packed in a T=1 icosahedral shell when the height of the apices is $0.5r$. The edge of the capsomer, a, which is the edge of a pentagon, is related to the radius through Eq.3.7

$$a = r\sqrt{\frac{5-\sqrt{5}}{2}} \tag{3.7}$$

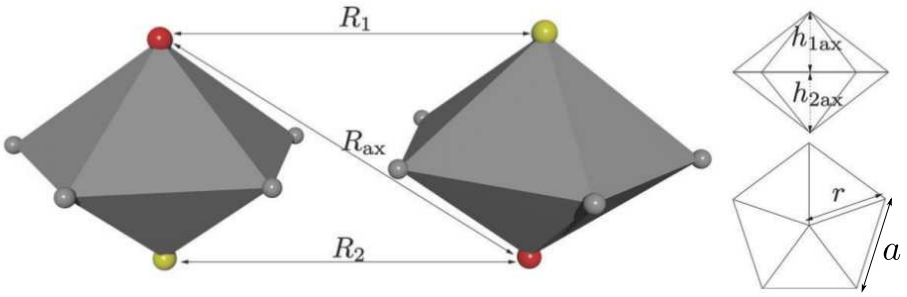

Figure 3.8: Left: Definition of the interactions between the apex sites of two pyramids. Primary apex sites are coloured in red. Rigth:geometrical properties of a pentagonal pyramid

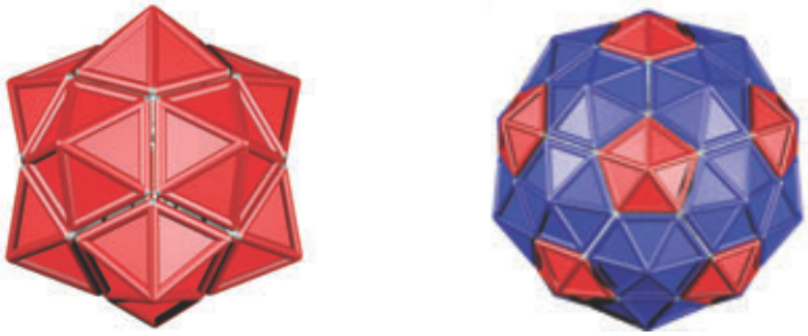

Figure 3.9: T=1 and T=3 capsids obtained with the updated Wales Model. Pentamers are colored in red and hexamers in blue. Figure adapted from [181]

Fejer et al [181] studied the energy landscape of the systems with triangulation numbers T=1 and T=3. T=1 capsids were assembled from 12 pentameric subunits and T=3 capsids contained twenty hexamers and twelve pentamers, and found that the global minimum exhibited correct shell formations for the right set of parameters (Fig. 3.9).

Johnston, Louis and Doye [182] performed MD simulations for the original Wales model [180] where the pyramids contain only one appex site at the top. They studied the assembly dependence with temperature and concentration for a given set of parameters.

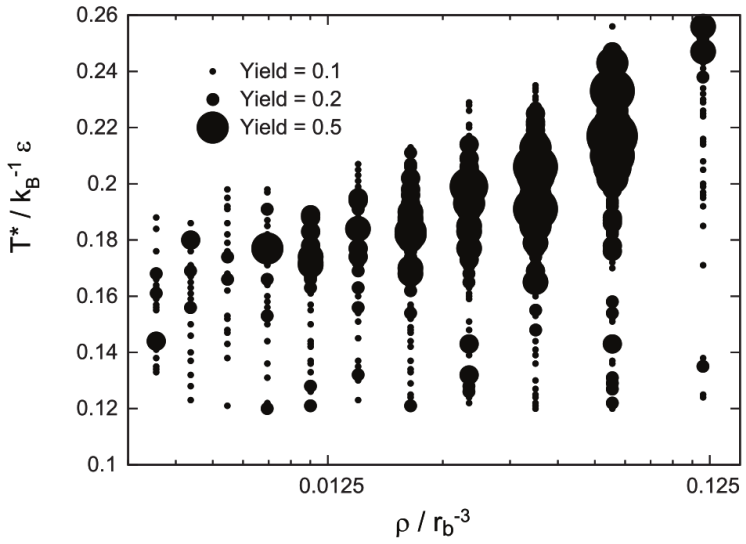

Figure 3.10: Assembly yield as a function of density temperature. The assembly yield is defined as the fraction of fully formed capsids. Figure adapted from [182]

They observed that for fix density, the region of optimal assembly was bounded from above and below in temperature (Fig. 3.10). If the temperature drops too low, then the attractions are too strong and mis-bonded capsomers cannot dissociate and reassemble in correct configurations. If the temperature is too high, the attractions are not strong enough to ensure bonding, and the high entropy disordered state was favored. Similarly, they found that at fixed temperature, there was also a window of densities for which optimal yields are obtained.

4 Sequential assembly and budding

4.1 Introduction

Assembly and budding do not occur simultaneously for most DNA viruses and some retroviruses; a capsid is usually formed prior to budding. From the capsid assembly to the virion release outside the cell, the final steps of the viral replication cycle can follow very different routes depending on the assembly sites, the different transport processes, and the diversity of proteins mediating the attachment to the membrane. Despite this great diversity, some general aspects are common in many virus families for which assembly and budding happen sequentially.

If we focus on these common features, the exit process can be described as the attachment of an already assembled capsid with spherical or cylindrical geometry to a membrane, followed by the wrapping of the particle until it is completely surrounded and only a thin neck keeps it connected to the original membrane. The fusion of the neck completes the process with the release of the enveloped virus to the extracellular space when budding takes place on the plasma membrane, or to the cytoplasm when budding occurs in an internal cellular membrane (Fig. 4.1).

These general aspects of budding are not only shared by virus families that follow this pathway to acquire the membrane envelope, but also by nanoparticles used in biomedical applications. The mechanisms by which viral and nanoscale particles cross cell membranes and the factors that control their uptake are essential questions for cellular physiology and modern biomedicine. Regulating the uptake (endocytosis) of nanoparticles is important for nanomedicine applications and for predicting nanoparticle toxicity [183–185].

Although endocytosis processes [186], viral budding [68, 187], and scission of budded

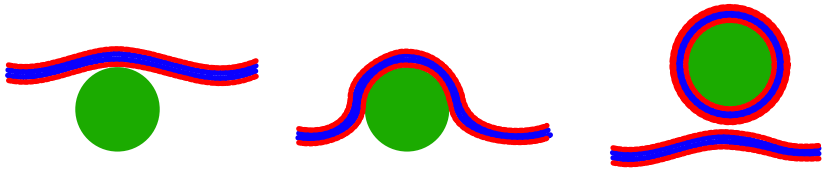

Figure 4.1: During budding, the particle first adheres to the membrane **(a)**, then is progressively wrapped **(b)**, and finally buds out of the membrane **(c)**.

viruses [188] can be actively driven or assisted by cell machinery, both nanoparticle uptake and at least some aspects of budding of viruses or viral proteins [189, 190] can occur passively (without cell machinery or ATP hydrolysis) [69, 75, 191, 192]. Evidence also suggests that some viruses do or can undergo passive budding in vivo (e.g. [192, 193]). It is therefore important to establish the aspects of particle budding which are generic to passive transport and thus underlie all forms of particle uptake or egress.

In this chapter, we aim to investigate the basic and common features to all these mechanisms of budding, using elastic theory, MD simulations, and free energy calculations to characterize the dynamics and thermodynamics of the budding process.

Previous works first studied the equilibrium configurations of budding through a vesicle or infinite membrane as a function of membrane rigidity, particle size, and membrane-particle adhesion energy using elasticity theory [194–197]. Subsequently, Monte Carlo simulations on a vesicle represented by a randomly triangulated surface [198] and MD on a coarse-grained lipid model [199] were used to investigate wrapping of charged particles, while density functional theory was applied to study the relationship between particle hydrophobicity and wrapping [200]. Dissipative particle dynamics was also used to investigate the wrapping of a particle by a inhomogeneous bilayer [201] and the wrapping behavior of ligand-coated nanoparticles [202]. Recently, the wrapping behavior of ellipsoidal particles has been studied via DPD [203] and MD simulations [204].

While all of these treatments show that the adhesion energy required for wrapping depends on particle properties and membrane composition, there has not been a thorough comparison of predictions of elasticity theory with the results of more sophisticated computational models. In this work our primary objective is to understand the extent to which simplified elastic models can describe the thermodynamics and/or dynamics of particle uptake, and to elucidate the basic aspects and minimum requirements for budding.

To focus on aspects generic to all forms of exo- or endocytosis, we consider a minimal model in which the membrane is treated as a bilayer of homogeneous composition and the spherical particle is pre-assembled. Thus, we will not consider the effects of membrane inhomogeneity (i.e. lipid rafts) or the association of viral membrane proteins. In this way, budding will be solely driven by a generic effective interaction between the surface of the spherical particle and the membrane. We will use the predictions of the elastic model proposed by Deserno [194] to compare with our results of dynamical simulations and free energy calculations

4.2 Methods

4.2.1 Membrane Model and Membrane-particle interaction

We represent the membrane using the Cooke Model described in Section 3.2.2 (Eqs. 3.1 - 3.4). As we stressed before, this model offers very good features for its use in minimal representations of coarse grained systems, due to its simplicity, tunability and well behavior. This model allows the formation of bilayers with physical properties such as fluidity, area per molecule and bending rigidity that are easily tuned via ω_c (Fig. 4.2). Moreover, diffusivity within the membrane, density, and bending rigidity are in good agreement with values of these parameters measured for biological membranes [93].

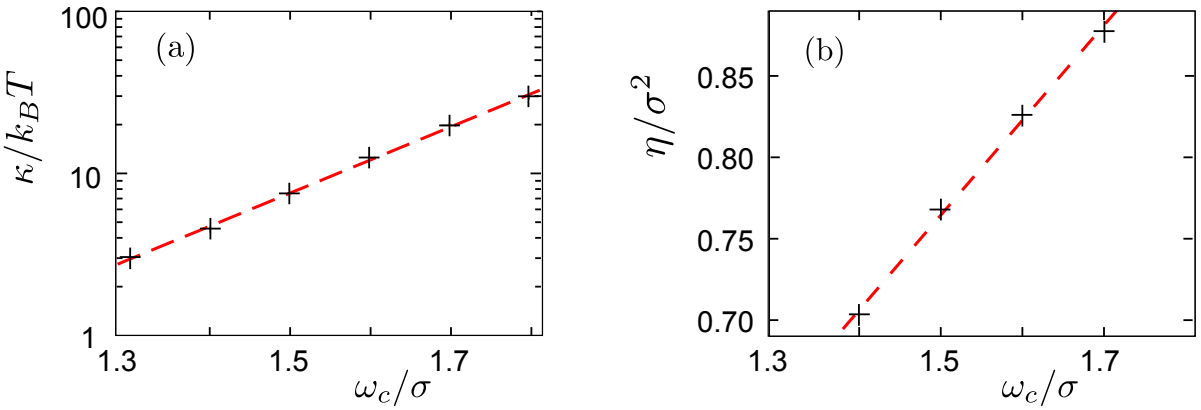

Figure 4.2: The properties of the membrane are easily tuned by ω_c, here the values for $k_BT = 1.1$ are showed.**(a)** Bending rigidity κ (in log scale) and **(b)** areal density η of lipids as functions of ω_c. The values in **(a)** are from Ref. [93] and the values in **(b)** were calculated from our simulations.

According to our system hypothesis, buding of virions or nanoparticles is driven by attractive interactions with lipid membranes. These interactions can arise in part from electrostatic potentials between charged lipid head groups and charges on the nanoparticle surface or capsid exterior (e.g. basic residues on the matrix protein in retroviruses [68]). A second source of interaction can be protein mediated, including binding of nanoparticle-functionalized ligands to membrane receptors or insertion of hydrophobic tails on capsid proteins into the membrane [205]. Finally, transmembrane viral spike proteins can drive or facilitate budding. Importantly, each of these forms of interactions is short ranged. Receptor-ligand and spike protein-virus interactions operate on length scales of Å to nm; similarly, at physiological conditions of 100 mM salt, electrostatic interactions have a Debye screening length of 1 nm. Thus, to keep our analysis general, we consider a short range attractive interaction between our model particle and head groups. In particular, we represent the combination of excluded volume and attractive interactions between the particle and head groups with a shifted Lennard

Jones potential:

$$V_{\text{particle-head}} = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r-s}\right)^{12} - \left(\frac{\sigma}{r-s}\right)^{6}\right] - V_{\text{cut}} & ; r < r_{\text{ph}} + s \\ 0 & ; r \geq r_{\text{ph}} + s \end{cases} \tag{4.1}$$

with ϵ being a free parameter that controls the membrane-particle interaction strength, $s = R - \sigma/2$, R the particle radius, $V_{\text{cut}} = 4\epsilon\left[\left(\frac{\sigma}{r_{\text{ph}}}\right)^{12} - \left(\frac{\sigma}{r_{\text{ph}}}\right)^{6}\right]$, $r_{\text{ph}} = 3.5\sigma$, and $\epsilon - V_{\text{cut}}$ the depth of the attractive interaction between the particle and the membrane. Notice that parameter s shifts the range of the Lennard-Jones interaction so that the membrane interacts with the particle surface instead of with its center.

The particle experiences only excluded volume interactions with the tail groups, which are modeled with a shifted WCA potential [155]:

$$V_{\text{particle-tail}} = 4\epsilon_0\left[\left(\frac{\sigma}{r-s}\right)^{12} - \left(\frac{\sigma}{r-s}\right)^{6} + \frac{1}{4}\right]. \tag{4.2}$$

4.2.2 Adhesion free energy

To enable comparison between the simulation results and the elastic theory presented below (Eq. 4.7), it is important to note that budding is controlled by particle-membrane adhesion free energy rather than simply the depth of the head group-particle attractive potential well ϵ. Specifically, the elastic theory will be presented in terms of ϵ^*, the adhesion free energy per area. We calculate this quantity from the simulated potential as

$$\epsilon^*/k_{\text{B}}T = -\eta \log\left[1 + \int_{s+\sigma}^{\infty} dr\left(e^{-V_{\text{particle-head}}(r)} - 1\right)\right]. \tag{4.3}$$

The resulting relationship is shown in Fig. 4.3. Here we have neglected the tiny contribution from cutting off the potential, we assume that each lipid head group in contact with the particle approaches roughly along a radial coordinate, and we neglect configurational entropy losses endured by the lipid molecules during adhesion.

4.2.3 Parameters

From the phase diagram in Ref. [93] we set the temperature of our simulations to $k_{\text{B}}T/\epsilon_0 = 1.1$, which allows for a broad range of ω_{c} (between 1.3σ and 1.7σ) within which the membrane is in the fluid state. Furthermore, the bending rigidity was calculated as a function of ω_{c} over this range of values for $k_B T/\epsilon_0 = 1.1$ in the same work; results from that reference are shown in Fig. 4.2a. Values of the bilayer density calculated in our simulations over a similar range of ω_{c} are shown in Fig. 4.2b.

The units of energy, length, and time in our simulations are respectively ϵ_0, σ and τ_0. The remaining parameters can be assigned physical values by setting the system to room

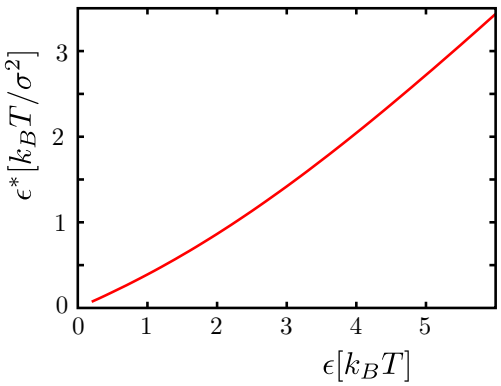

Figure 4.3: Relationship between the adhesion energy, ϵ, and the adhesion free energy per area, $\epsilon*$, for $\kappa = 13.9 k_B T$.

temperature, $T = 300K$, and noting that the typical width of a lipid bilayer is around 5 nm, and the mass of a typical phospholipid is about 660 g/mol. The units of our system can then be assigned as follows: $\sigma = 0.9$ nm, $m_0 = 220$ g/mol, $\epsilon_0 = 3.77 \times 10^{-21}$J $= 227$g$\text{Å}^2/\text{ps}^2$mol, and $\tau_0 = \sigma\sqrt{m_0/\epsilon} = 8.86$ ps.

4.3 Simulations

Molecular dynamics (MD) simulations of budding were performed at constant temperature and tension using the velocity Verlet algorithm, with a Langevin thermostat [111] to maintain constant temperature and a modified Andersen barostat [120] to simulate a tensionless infinite membrane. The algorithm used is the one presented in Section 2.3.3 with Eqs. 2.39 to 2.41. This statistical ensemble reproduces the conditions of the cellular environment, where temperature is constant around 37 °C and the membrane is tensionless on average. The time step was $\Delta t = 0.01\tau_0$, the friction constant was $\gamma = \tau_0^{-1}$, the box friction for the Andersen barostat was $\gamma_{\text{box}} = 2 \cdot 10^{-4}$ and the box mass $Q = 10^{-5}$ in the system units. The reference pressure, P_0, was set to 0, to simulate a tensionless membrane. The tension equals the pressure because the normal component to the membrane, the z-axis in our case, is free to fluctuate and does not contribute to the pressure. The x and y components of velocities and positions were rescaled according to the changes in the volume. In order to simulate an infinite membrane, periodic boundary conditions were employed.

For most simulations the membrane was comprised of $n = 7,164$ lipids. An initial bilayer configuration was relaxed by MD and then placed normal to the z-axis in a cubic box of side-length $L = 63.5\sigma$. The particle was introduced in the center of the box with its top pole located about 5σ below the membrane surface with zero initial velocity.

Since the membrane was kept tensionless by the barostat, the size of the box decreased during simulations as the particle was wrapped. To ensure that there were no finite size effects,

additional sets of simulations were performed, following the same protocol, for membranes with $n = 16,200$ lipids in an initial box size of 100x100x60 σ^3, and with $n = 28,800$ lipids in an initial box size of 130x130x60 σ^3. Except where mentioned otherwise, results are shown for the system with $n = 7,164$ lipids.

Free energy calculations. In addition to performing dynamical simulations of budding, we calculated the potential of mean force as a function of particle penetration using umbrella sampling (Section 2.3.4). Simulations were performed in which the system was biased toward particular values of the penetration p by introducing a biasing function $U_{\text{bias}}(\{\mathbf{r}\}) = \frac{1}{2}\kappa_{\text{umb}}(p(\{\mathbf{r}\}) - p_0)^2$. Here $p(\{\mathbf{r}\})$ is the penetration for a configuration $\{\mathbf{r}\}$ and is defined as the distance between the top of the particle and the center of mass of the membrane. A series of windows were performed at different values of p_0; for all windows $\kappa_{\text{umb}} = 200\epsilon_0/\sigma^2$. The simulations were started for an unwrapped particle ($p = -6\sigma$), and initial coordinates for each subsequent window were obtained from simulations in the previous one. Statistics from each window were stitched together and re-weighted to obtain the unbiased free energy using the implementation of WHAM (Section 2.3.4) from the Grossfield Lab [206].

An additional set of free energy calculations was performed for $\epsilon = \epsilon_0$ and $R = 12\sigma$ (which lead to metastable partially wrapped states as described below) to improve statistics and to determine if the harmonic potentials were constraining relaxation of other degrees of freedom. A bias potential given by the negative of the already calculated free energy was imposed. Sampling was performed in a series of windows, wherein a smoothed Heaviside function (Eq. 4.4) constrained the penetration around around a window-dependent value of p_0 within a range of 0.4σ:

$$U_{\text{Heaviside}} = \begin{cases} \frac{1}{2}\kappa_{\text{umb}}(p(\{\mathbf{r}\}) - p_{\max})^2 & ; p \geq p_{\max} \\ \frac{1}{2}\kappa_{\text{umb}}(p_{\min} - p(\{\mathbf{r}\}))^2 & ; p \leq p_{\min} \end{cases} \tag{4.4}$$

where $p_{\max} = p_0 + 0.2\sigma$ and $p_{\min} = p_0 - 0.2\sigma$

A third set of free energy calculations was performed in which only smoothed Heaviside constraints were applied. Similar unbiased free energies were obtained in each case, indicating that relaxation of other degrees of freedom was satisfactory. Histograms from all sets of free energy calculations were then stitched together using WHAM.

4.3.1 Elastic model

To evaluate the results of the dynamical simulations and free energy calculations, we compare the simulation results to a simplified elastic model for invagination of a particle in a membrane. Our elastic model closely follows that of Ref. [194] but we consider an infinite tensionless membrane rather than a vesicle.

The total energy of the particle-membrane system arises from the energy of adhesion between the particle and the membrane (e_{ad}) and the elastic energy of the membrane (e_{m}).

Following the simulation model, we assume that adhesion is mediated by short-range interactions with energy per area $-\epsilon^*$ so that the total energy of adhesion is $e_{\text{ad}} = -\epsilon^* a_{\text{wrap}}$ with a_{wrap} the area of the membrane in contact with the particle. Note that ϵ^* actually describes a free energy since it includes the effects of counterion dissociation and other entropic factors involved in particle associations, but following Ref. [194] we refer to it and the elastic terms described next as energies to emphasize that we are neglecting the (small) contribution to the free energy associated with fluctuations around the lowest free energy membrane configuration.

To calculate the elastic contributions to the energy, we consider the Helfrich Hamiltonian for an infinitesimally thin membrane [130]

$$e_m = \int da \left(\sigma_s + \frac{\kappa}{2}(2H - C_0)^2 + \kappa_{\text{G}} K\right) \tag{4.5}$$

where σ_s is the surface tension, κ and κ_G are the bending rigidity and the Gaussian curvature modulus respectively, H and K are the mean and Gaussian curvatures, and C_0 is the spontaneous curvature. Our model membrane is symmetric and tensionless, so C_0 and σ_s are 0. We will use this elastic model to describe the budding process up until the point of scission at the neck, and thus the topology of the membrane remains constant. Assuming that the Gaussian curvature modulus κ_{G} is invariant throughout the membrane, the last term in Eq. 4.5 is constant under the Gauss-Bonnet theorem [131]. The elastic energy for a general configuration of the membrane is then given by

$$e_{\text{m}} = \int \frac{\kappa}{2} \left(\frac{1}{r_1} + \frac{1}{r_2}\right)^2 da \tag{4.6}$$

with r_1 and r_2 the principal radii of curvature.

We follow Deserno et al. [194] to assume that the following geometry closely corresponds to the lowest free energy configuration for a partially wrapped particle (Fig. 4.4). There is an area a_{wrap} of the membrane tightly adhered to the particle with radius of curvature approximately equal to the particle radius R, and a rim area, a_{rim}, between the point at which the membrane separates from the particle and where it recovers a flat configuration. Because the particle is a featureless sphere, we assume that the lowest free energy configuration is axisymmetric, to give the energy

$$e = e_{\text{ad}} + e_{\text{wrap}} + e_{\text{rim}} = -a_{\text{wrap}}\epsilon^* + \kappa \left[\frac{2a_{\text{wrap}}}{R^2} + \int \frac{da_{\text{rim}}}{2} \left(\frac{1}{r_1} + \frac{1}{r_2}\right)^2\right] \tag{4.7}$$

where a_{rim} is the area of the rim surrounding the particle, and r_1 and r_2 are the principal radii of curvature in the rim area.

We now recast Eq. 4.7 in terms of two new variables, the latitudinal degree of wrapping θ and the penetration p, which is the distance the particle travels along the direction normal to the flat membrane, measured from the point at which the surface of the particle first touches

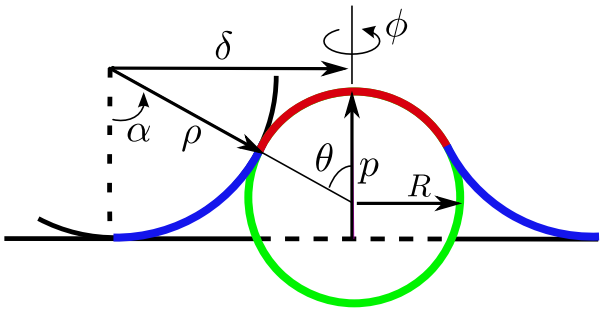

Figure 4.4: Cross-section of the 3D geometry used for the elastic model of a membrane wrapping a particle of radius R. The particle, depicted as a green sphere, adheres to a section of the membrane in red with area a_{wrap}. The surrounding membrane, with area a_{rim}, drawn in blue, decays toward the flat configuration. The shape of this surrounding membrane is taken to be a section of a torus for simplicity. δ and ρ stand for the outer and inner radius of the section of the torus formed by the rim region, and α and ϕ represent the polar and azimuthal angle of spherical coordinates. For a given penetration, p, there is a wrapping degree, θ, that minimizes the elastic energy.

the flat membrane. The maximum possible degree of wrapping for a given penetration is thus defined by $\theta_{\text{max}} = \text{acos}(1 - p/R)$. Note that while the theoretical penetration describes the vertical distance between the particle top and the position of the flat membrane, the computational definition describes the vertical distance between the particle top and the center of mass of the membrane. The two definitions are thus qualitatively the same but not identical.

In Fig. 4.4, a schematic of the system is depicted as a 2-D cross-sectional cut, in which the red line represents the section of the membrane bound to the particle and the blue line represents the rim region. As shown in the schematic, we assume that the rim corresponds to a section of a torus (appearing as a circular arc in the 2-D cross-section). Although this is only one of the multiple shapes the rim can form, it was shown to closely correspond to solutions from a full variational calculation in Ref. [194] and allows us to write the geometric properties of the system as explicit functions of our parameters. In particular, the radius of the torus depends uniquely on the particle size R, the wrapping degree θ and the penetration p; the area element on a torus and the two principal radii of curvature are [194] $\text{d}a_{\text{rim}} = \rho(\delta - \rho\sin\alpha)d\alpha d\phi$, $r_1 = \rho$, and $r_2 = -\frac{\delta - \rho\sin\alpha}{\sin\alpha}$, with α and ϕ as the polar and azimuthal angles in spherical coordinates. With the new parametrization, the area of the membrane in contact with the particle for a wrapping degree θ turns out to be $a_{\text{wrap}}(\theta) = 2\pi R^2(1 - \cos(\theta))$.

Therefore, the energy of the system can be written in the following way:

$$e = \left(-\frac{A\epsilon^*}{2} + 4\pi\kappa\right)(1-\cos\theta) \tag{4.8}$$

$$+ \pi\kappa \int_0^\theta \rho|\delta - \rho\sin\alpha| \left(\frac{1}{\rho} - \frac{\sin\alpha}{\delta - \rho\sin\alpha}\right)^2 d\alpha \tag{4.9}$$

$$\rho = \frac{p + R(cos\theta - 1)}{2sin^2(\theta/2)} \tag{4.10}$$

$$\delta = -\frac{p}{tan(\theta/2)} \tag{4.11}$$

where $\theta \in (0, \theta_{\max})$, $\theta_{\max} = \text{acos}(1 - p/R)$ and $A = 4\pi R^2$ is the surface area of the particle. For a given bending rigidity, particle size and membrane-particle interaction, the energy only depends on the penetration p and the wrapping degree θ.

4.4 Results

4.4.1 Elastic theory phase diagram

For each value of p we minimize the energy (Eq. 4.11) with respect to θ to obtain the membrane configuration and corresponding energy as a function of penetration alone. The results of the minimization are shown in Fig. 4.5. Here, the dashed line is the binodal, given by $\epsilon* = \frac{2\kappa}{R^2}$, below which wrapping is energetically unfavorable, and the solid line denotes the spinodal, above which wrapping proceeds without any energetic barrier. In between the lines there is a barrier to wrapping which begins at $p = 0$, meaning that there are no long-lived partially wrapped states consistent with this theory for infinite membranes.

4.4.2 Simulation results

To understand the influence of membrane and particle properties on budding, we began by performing dynamical simulations for a range of particle-membrane interaction strengths, ϵ, particle radius R, and ω_c, which controls the areal density of lipids, the bending rigidity κ, and diffusion rates within the membrane, as described in section **The membrane model**. Different values of these parameters lead to dramatically different behaviors, as shown in the phase diagrams presented in Fig. 4.6. First we note that the behaviors can be grouped into four classes, which we illustrate by describing trajectories observed for various values of ϵ and constant values of $\kappa = 13.9 k_B T$ and $R = 12\sigma$ (Fig. 4.7).

For weak adhesion strengths ϵ, no wrapping occurs; the membrane continues to exhibit only the usual spectrum of thermal fluctuations (Fig. 4.8) after the particle adheres to it, and the penetration oscillates around negative values (Fig. 4.7, case for $\epsilon = 0.5\epsilon_0$). For a narrow intermediate range of $0.6 < \epsilon/\epsilon_0 \lesssim 1.2$, the particle adheres to the membrane, but wrapping ceases at a partially wrapped state (Fig. 4.9), after which the degree of particle

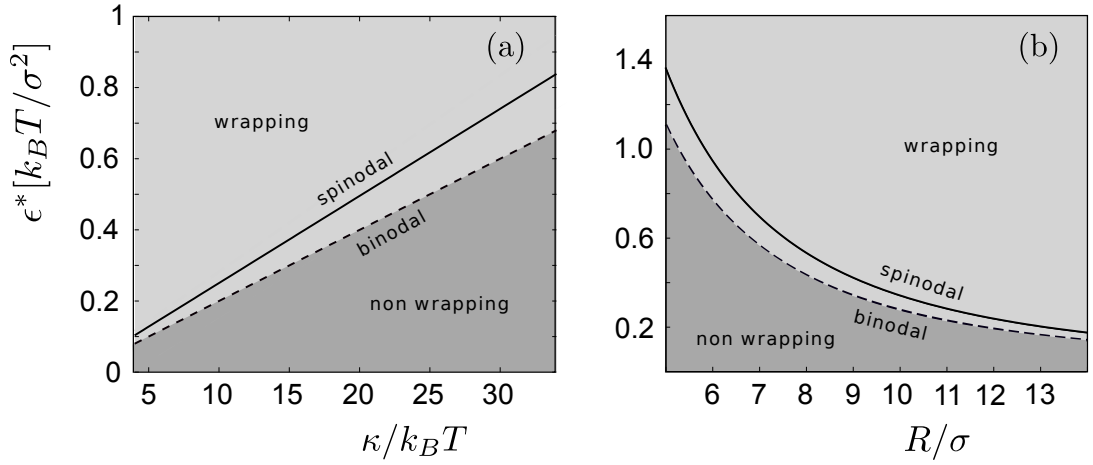

Figure 4.5: Phases diagrams obtained from the energy minimization in Eq. 4.11 (a) as a function of ϵ^* and κ for constant $R = 10\sigma$ and (b) as a function of the particle size, R, and the adhesion free energy per area, ϵ^*, for constant bending rigidity $\kappa = 13.9k_BT$. The binodal ($\epsilon* = \frac{2\kappa}{R^2}$) above which wrapping is energetically favorable is indicated by a dashed line and the spinodal, above which wrapping proceeds without an energy barrier, is shown by a solid line.

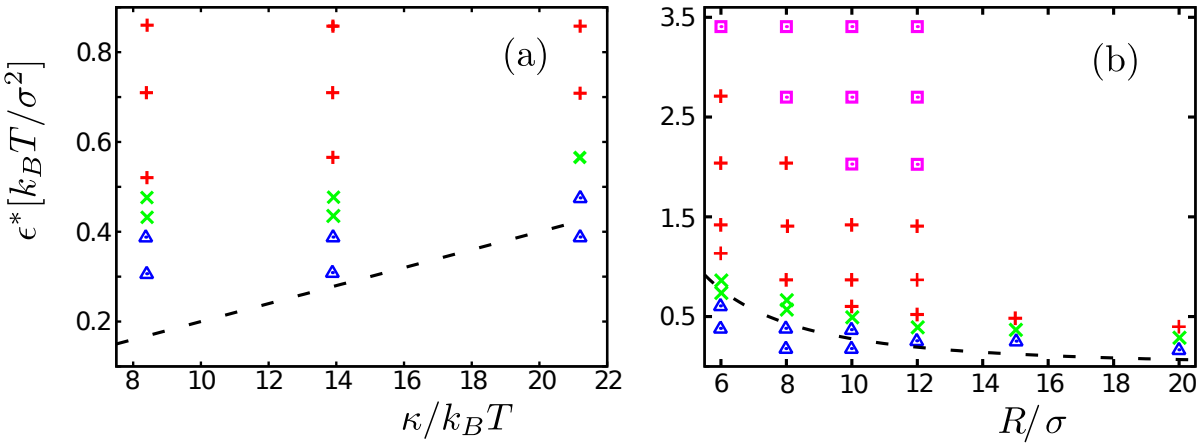

Figure 4.6: Phase diagrams obtained from MD simulations and free energy calculations (a) as a function of ϵ and κ for constant $R = 10\sigma$ and (b) as a function of the particle size R and adhesion free energy $\epsilon*$ for constant bending rigidity, $\kappa = 13.9k_BT$. Parameter sets are identified as those which lead to no wrapping (△ symbols), long-lived partially wrapped structures (x symbols), complete wrapping (+ symbols), and those for which the membrane undergoes rupture prior to budding (□ symbols). The wrapping binodal predicted by the elastic theory is shown as a dashed line on each plot. The relationship between the adhesion free energy $\epsilon*$ and the head group-particle attractive well depth ϵ is given in Eq. 4.3 and in Fig. 4.3.

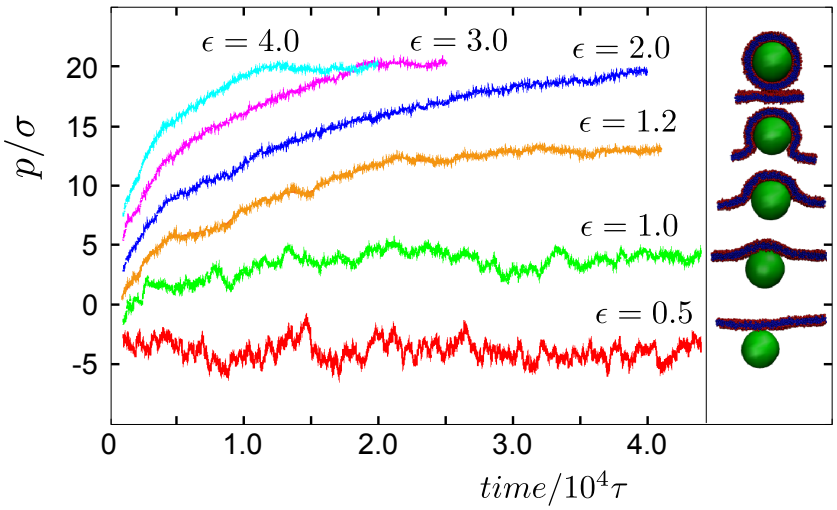

Figure 4.7: Particle penetration into the membrane, p, as a function of time for MD trajectories with different values of the adhesion energy ϵ for a system with particle radius $R = 12\sigma$ and membrane bending rigidity $\kappa = 13.9k_BT$. For $\epsilon = 0.5\epsilon_0$ (red line) no wrapping occurs. For $\epsilon = 1.0\epsilon_0$ (green line) and $\epsilon = 1.2\epsilon_0$ (orange line), budding becomes stalled at a partially wrapped state whose value increases with ϵ. For $\epsilon = 2.0\epsilon_0$ (blue line), 3.0 (pink line), and 4.0 (cyan line) the particle undergoes complete encapsulation. On the right, slices of the system for corresponding values of p are shown. Images were generated using VMD [207].

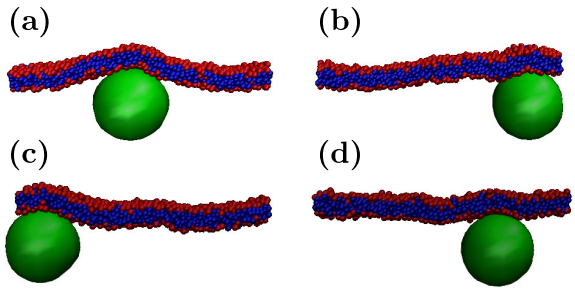

Figure 4.8: Adhesion without wrapping. Slices of configurations extracted from MD simulations with $\epsilon = \epsilon_0$, $R = 10\sigma$, and $\kappa = 13.9k_BT$. Times shown are (a) $t = 5 \cdot 10^3\tau_0$, (b) $t = 4 \cdot 10^4\tau_0$, (c) $t = 6 \cdot 10^4\tau_0$, (d) $t = 1 \cdot 10^5\tau_0$.

penetration into the membrane fluctuates around a steady value (Fig. 4.7, cases for $\epsilon = 1.0\epsilon_0$ and $\epsilon = 1.2\epsilon_0$). The average value of the penetration remained unchanged for as long as we simulated (up to $4 \cdot 10^4\tau_0$). The final degree of penetration in this arrested state increases with ϵ, until approximately the point at which the particle is half wrapped (Fig. 4.7, case for $\epsilon = 1.2\epsilon_0$).

A further increase in ϵ results in the next class of trajectories (Fig. 4.7, case for $\epsilon = 2.0\epsilon_0$ and $\epsilon = 3.0\epsilon_0$), in which the particle is completely encapsulated (Fig. 4.10). In this case, wrap-

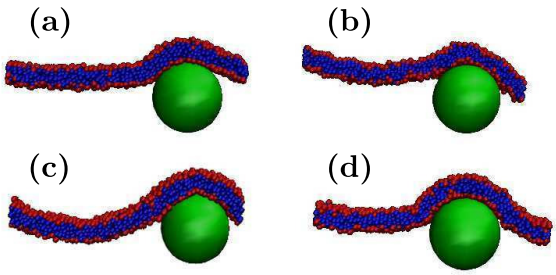

Figure 4.9: Long-lived partial wrapping. Slices of configurations extracted from MD simulations with $\epsilon = 1.3\epsilon_0$, $R = 10\sigma$, and $\kappa = 13.9k_BT$. The particle remains partially wrapped for the length of the simulation ($t = 4 \cdot 10^4\tau_0$). (a) $t = 5 \cdot 10^3\tau_0$, (b) $t = 1.5 \cdot 10^4\tau_0$, (c) $t = 2.5 \cdot 10^4\tau_0$, (d) $t = 3 \cdot 10^4\tau_0$.

ping proceeds steadily until the particle is completely surrounded by membrane except for a narrow neck region (Fig. 4.10d). Wrapping is then completed when a thermal fluctuation causes the neck to break and have its sides fused (Fig. 4.10e), after which the fully wrapped particle diffuses away from the membrane (Fig. 4.10f). Since fusion is a stochastic event, the budding time can be variable and we have observed neck configurations lasting between $500\tau_0$ and $5000\tau_0$. The elastic theory predicts that the shape and length of the neck depend on the balance between the adhesion energy and the bending energy, with strong adhesion favoring a short neck and large bending energies favoring a long neck. The simulation results are consistent with this prediction; example configurations are shown in Fig. 4.11. The figure shows snapshots from simulations with particle radius $R = 6\sigma$, bending rigidity $\kappa = 13.9k_BT$ and different values of ϵ. A small particle size was chosen for the figure because the relationship between neck configuration and adhesion energy is most easily visualized when high membrane curvature is required for wrapping.

The fact that fusion is accessible within the course of a typical simulation is an interesting contrast between the model studied here and that studied by Smith et al. [201], where fusion was observed only for inhomogeneous membranes, where the line tension arising from phase separated domains with an interface surrounding the particle was needed to promote fusion. The difference in fusion behavior between the two models could arise from the fact that Smith et al. [201] consider lipid molecules with two tails, but other parameter differences (e.g. range of bending moduli considered or lipid cohesion) might also play a role.

For higher values of ϵ wrapping proceeds extremely rapidly (Fig. 4.7, case for $\epsilon = 4.0\epsilon_0$) as there is a strong driving force to increase the number of head-particle interactions (Fig. 4.12). As the curvature of the membrane in the vicinity of the wrapping front increases, the membrane structure undergoes ruptures in that region (Fig. 4.12d), and a pore forms in the membrane (Fig. 4.12e). The pore size increases with the adhesion energy, with a maximum diameter of the order of the particle radius. The fully encapsulated particle then diffuses

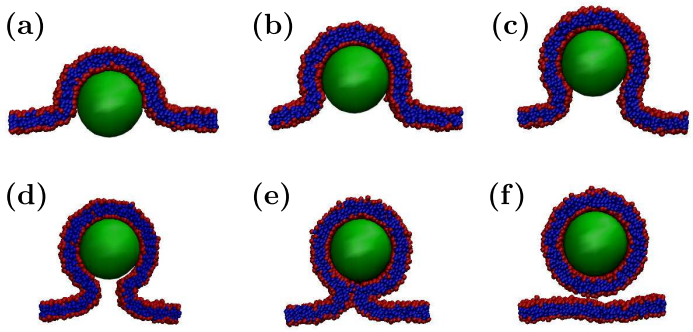

Figure 4.10: Wrapping. Slices of configurations at different times extracted from MD simulations with $\epsilon = 3\epsilon_0, R = 10\sigma$ and $\kappa = 13.9k_BT$. The membrane wraps the particle (a-d) until a neck or channel connecting the flat bilayer and the membrane surrounding the particle forms (e). Thermal fluctuations break this narrow neck, resulting in the encapsulated particle escaping from the membrane (f). Configurations are shown for times (a) $t = 5 \cdot 10^3\tau_0$, (b) $t = 1 \cdot 10^4\tau_0$, (c) $t = 1.5 \cdot 10^4\tau_0$, (d) $t = 1.95 \cdot 10^4\tau_0$, (e) $t = 2 \cdot 10^4\tau_0$, (f) $t = 2.5 \cdot 10^4\tau_0$.

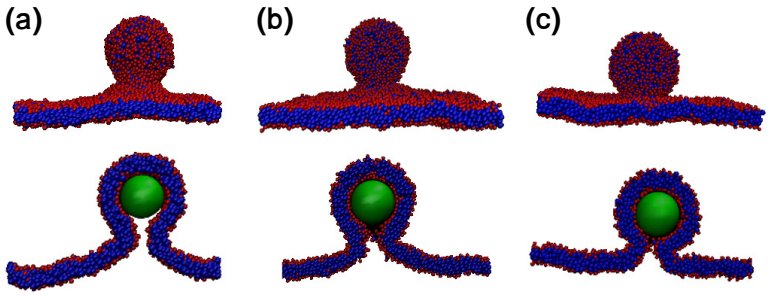

Figure 4.11: The neck profile depends on the adhesion strength. Membrane configurations are shown shortly before the completion of budding for $R = 6\sigma$, $\kappa = 13.9k_BT$ and different adhesion strengths. (a) A relatively small adhesion strength, $\epsilon = 3\epsilon_0$, leads to a long neck. (b) For $\epsilon = 4\epsilon_0$ the neck is shorter. (c) For $\epsilon = 5\epsilon_0$, close to the adhesion strength that leads to membrane rupture, the neck length is comparable to the height of typical membrane fluctuations. The top row of images shows a side view of system configurations and the bottom row of images gives the corresponding side-view slices.

away and the pore heals through thermal motions of the lipids. The formation of a pore during these budding trajectories resembles the process by which a hydrophobic nanoparticle passes through membranes in the simulations described in [202, 208], but the physical driving forces are different in this case and the pore arises for kinetic reasons. Namely, the collective wrapping process proceeds more slowly than ruptures form in the membrane due to the large driving force to increase particle-head group contacts.

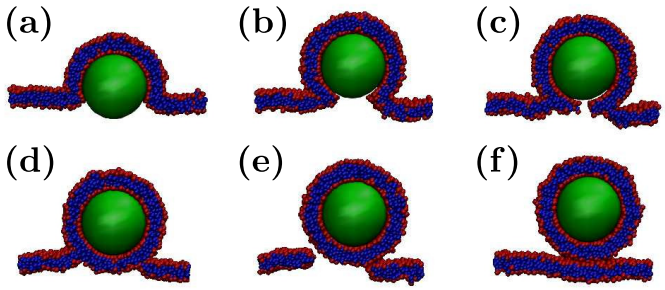

Figure 4.12: Wrapping via membrane rupture. Slices of configurations at different times extracted from MD simulations with $\epsilon = 5\epsilon_0$, $R = 10\sigma$, and $\kappa = 13.9k_BT$. Particle wrapping (a, b) leads to the formation of a pore (c, d). Eventually, the enveloped particle leaves the membrane (e) and the pore closes (f). Configurations shown occured at times (a) $t = 2 \cdot 10^3\tau_0$, (b) $t = 5 \cdot 10^3\tau_0$, (c) $t = 5.4 \cdot 10^3\tau_0$, (d) $t = 5.6 \cdot 10^3\tau_0$, (e) $t = 8 \cdot 10^3\tau_0$, (f) $t = 9.5 \cdot 10^3\tau_0$.

4.4.3 Free energy calculations and the partially wrapped state

We were particularly interested in the partially wrapped states seen in the dynamical simulations, as the elastic model predicts no stable or metastable partially wrapped states. To determine whether or not these observations corresponded to equilibrium configurations, the free energy was calculated as a function of the penetration using umbrella sampling (section **Simulations**). Calculated free energy projections are shown for three values of ϵ in Fig. 4.13, for which the finite-time dynamical simulations respectively ended in no wrapping ($\epsilon = 0.5\epsilon_0$), partial wrapping ($\epsilon = \epsilon_0$), and complete wrapping ($\epsilon = 1.25\epsilon_0$). For the cases of full wrapping and no wrapping, the calculated free energy projections are consistent with the dynamics results. Namely, for $\epsilon = 0.5\epsilon_0$ the minimum free energy value corresponds to no wrapping with a steep penalty for increasing penetration, while for $\epsilon = 1.25\epsilon_0$ the free energy decreases monotonically with increasing penetration until the particle is completely wrapped.

In contrast, the minimum value in the free energy profile for $\epsilon = 1.0\epsilon_0$ does not correspond to the partially wrapped state observed in the dynamical simulations, but rather corresponds to complete wrapping. The free energy profile does show a local minimum at $p \sim 5\sigma$, which corresponds to the penetration of the partial wrapping configuration, indicating that the states observed in the dynamical trajectories are metastable. However, the well depth is only a few k_BT while partially wrapped states continue for as long as we have simulated $4 \times 10^4\tau_0$. We therefore performed further comparison of umbrella sampling results to dynamical trajectories. We chose a set of configurations from the umbrella sampling trajectories with different values of p. For each such configuration we performed several unbiased MD trajectories initialized with velocities using different random number seeds

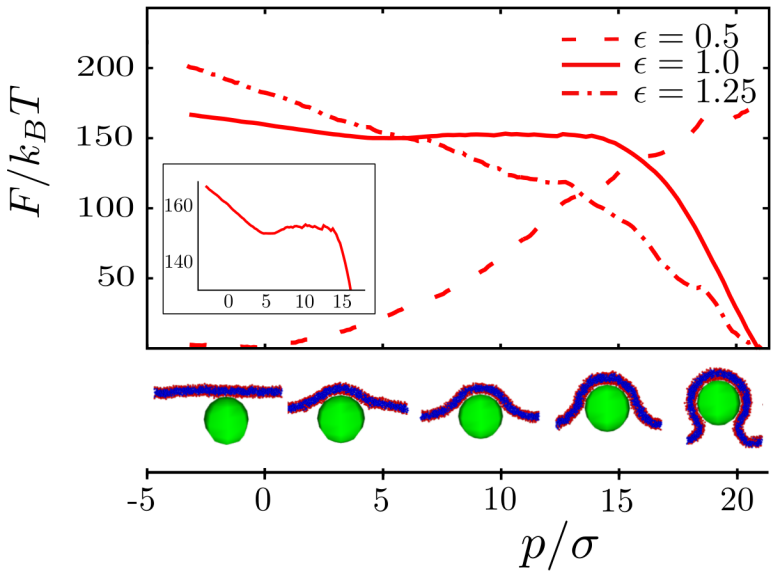

Figure 4.13: Free energy profiles as a function of the penetration, p, calculated from MD simulations using umbrella sampling are shown for $R = 12\sigma$, $\kappa = 13.9k_BT$, and indicated values of ϵ. In the inset, the local minimum for $\epsilon = 1.0$ can be observed. At the bottom, slices of the system as a function of the penetration are shown.

to obtain a crude estimate of the commitment probability [209]. We found that trajectories initiated from configurations with small values of $p \lesssim 5\sigma$ fluctuate around that value and configurations with $p \gtrsim 18.8\sigma$ progressed steadily to complete wrapping. However, configurations with $5\sigma \leq p \lesssim 18.8\sigma$ tended to fluctuate around the value of p corresponding to their initial configuration, which is inconsistent with the free energy profile for $p \geq 15\sigma$ and raises the possibility of an additional slow degree of freedom at moderate ϵ. I.e. while p is a suitable order parameter for determining the free energies of the stable states, it may not be a complete reaction coordinate capable of describing the dynamics along the transition [209].

It is not necessary to identify a perfect reaction coordinate to fulfill our primary objective of understanding the phase behavior, but we did examine other potentially relevant dynamical degrees of freedom. Analysis of umbrella sampling configurations during the equilibration phase of the calculation indicates that, when the particle is held at a fixed penetration, the membrane configuration gradually relaxes to a state of increased adhesion (Fig. 4.14). Thus a reaction coordinate capable of completely describing the transition dynamics may need to include an additional collective variable that describes adhesion and/or membrane bending. Investigating this possibility, however, is beyond the scope of the present work focused on the phase behavior.

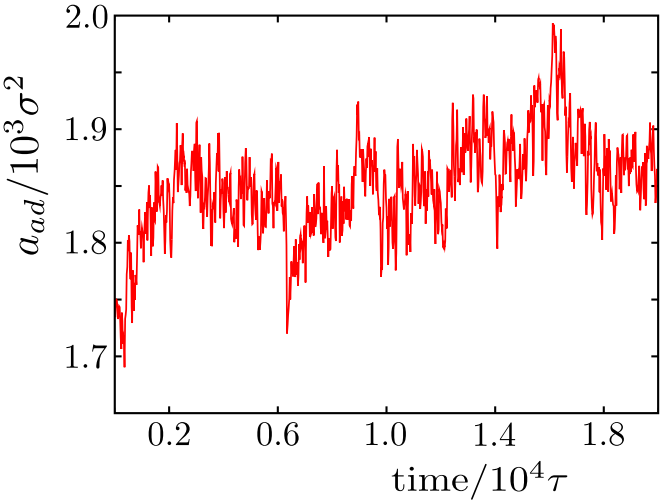

Figure 4.14: Evolution of the membrane area in contact with the particle as a function of time for MD simulations in which a smoothed Heaviside biasing potential (Eq. 4.4) holds the penetration in the vicinity of $p = p_0 = 16.875\sigma$. The time courses are averaged over 20 independent trajectories.

4.4.4 Comparison between computational and theoretical phase diagrams

Based on the results of dynamical simulations over a wide range of parameters, as well as umbrella sampling at parameter sets near the transition between no wrapping and wrapping, we determined phase diagrams as functions of adhesion energy ϵ, membrane bending modulus κ, and particle radius R (Fig. 4.6). To enable comparison with the elastic theory (Eq. 4.7, Fig. 4.5), it is essential to note the the theoretical parameter ϵ^* corresponds to the adhesion free energy rather than simply the depth of the head group-particle attractive potential well ϵ (see **Methods** section). Therefore, we plot the data as a function of the adhesion free energy per area ϵ^*; the relationship between ϵ^* and ϵ is shown in Fig. 4.3. The parameter sets for which non-wrapping is the equilibrium configuration are shown with $\triangle$ symbols, while the parameter sets which lead to equilibrium wrapping are separated into those which involve long-lived partially wrapped structures (x symbols), complete wrapping (+ symbols), and those for which the membrane undergoes rupture prior to budding ($\square$ symbols). The wrapping binodal predicted by the elastic theory is shown as a dashed line on each plot. We see that while the theory and simulations agree to within about $0.2k_BT$, the theoretical binodal is below the computational results. This discrepancy could occur because we have not accounted for the configurational entropy contributed by the lipids during adhesion or due to the fact that the theory assumes an infinitesimally thin membrane. A more significant difference between the computational results and the elastic theory is that, as noted above, the theory does not predict the existence of the metastable partially wrapped states.

4.5 Discussion

Our dynamical simulations of a minimal molecular model for the process of passive endo- or exocytosis identified four classes of behaviors resulting from the interaction of a particle with a membrane: no wrapping or adhesion resulting in a minimal perturbation of membrane configurations, partial wrapping, complete wrapping, and wrapping via rupture of the membrane. Equilibrium calculations showed that there are only two equilibrium configurations, corresponding to no wrapping or complete wrapping, and this equilibrium phase behavior for the molecular model shows strong agreement with the predictions of a simplified elastic theory [194]. The primary difference between the elastic model and the finite-time dynamical simulation results is the appearance of long-lived partial wrapping states.

Since the long-lived partially wrapped states seen in this study could be significant for dynamical, time-sensitive particle uptake processes such as endocytosis or viral budding in living organisms, it is worth comparing them to observations of other models. Most closely related to our results, Yue and Zhang [202] study a model comparable to ours except that the particle is coated with ligands that experience attractive interactions with membrane lipids. It appears that some of the configurations which they denote as 'adhesion' correspond to our long-lived partially wrapped states; however, it would be necessary to perform free energy calculations to determine whether they are metastable as we find here. In contrast to their model, we do not identify any parameter sets for which the particle partially penetrates into the hydrophobic interior of the membrane. Partially wrapped states have also been predicted from equilibrium theories in the context of finite-sized systems. Deserno et al. [194] examined the budding of a colloid from the interior of a spherical vesicle using the same Hamiltonian as introduced in elastic theory described above. They found that partial wrapping corresponds to the equilibrium state when the vesicle size is on the order of the colloid diameter due to the increase in curvature energy of the finite-sized vesicle. We similarly obtain partial wrapping configurations as equilibrium solutions if we introduce finite size into the elastic theory studied here, Eq. 4.7, by minimizing the energy of the elastic theory, Eq. 4.11, with the total membrane surface area constrained to $16R^2$, so that the wrapping area plus the rim area cannot exceed the total area. Zhang and Nguyen [196] also identify partially wrapped states as equilibrium solutions to an elastic theory, but they observed that the catenoid configuration is the only solution to the full variational problem for a tensionless infinite membrane. This solution implies that the elastic energy of the rim is always zero, and thus wrapping is only determined by the balance between the bending energy in the wrapped region and the adhesion energy, which does not lead to partial wrapping states. Because the toroid approximation for membrane configurations assumed in our elastic model is more restrictive than the full variational problem, the wrapping binodal shown in Fig. 4.5 is shifted to slightly higher values of the adhesion energy ϵ than obtained for their theory [196], but the behavior is qualitatively unchanged. Importantly, neither theory predicts the

partially wrapped state as a metastable configuration in an infinite membrane.

The membrane size in our simulations was chosen large enough to ensure that the theoretically predicted finite-size effects would not affect our results. To confirm that this was the case, we ran additional dynamical simulations and umbrella sampling calculations with membranes which were 50% and 100% larger (16200 and 28800 lipids respectively). The simulation results were the same for all three membrane sizes.

Finally, we consider our minimal model for passive endo- or exocytosis in the context of physical systems. Based on the length scales discussed in section **Parameters**, the particle diameters in our simulations range from 9 to 36 nm. Nanoparticles are available in a wide range of sizes, with particles smaller than 50 nm undergoing the most efficient uptake [210], although interaction between nanoparticles can reduce that radius when smaller nanoparticles cluster and bud together [197]. Our simulated particles are somewhat smaller than the size of viral capsids that undergo budding, which range from about 40 nm (e.g. hepadnavirus [211]) to hundreds of nanometers, but the results can be extrapolated into that range. As shown in Fig. 4.5 the adhesion-wrapping transition decreases with radius as $1/R$ for constant bending rigidity.

The limit of a tensionless membrane studied here could become a poor approximation if adhesion to the budding particle leads to crystalline order, but we did not see evidence of that in our simulations even at the highest adhesion energies studied. Other cases in which a tensionless membrane limit might break down include multicomponent membranes with lipids of different chain lengths or phase-separated regions leading to line tensions and thus spontaneous curvature [212]. In those cases, the range of adhesion energies that allow budding would be shifted according to the gain or loss in bending energy from the spontaneos curvature.

Our simulation results indicate that the existence of attractive interactions between a particle and lipid head groups, which has been proposed as the minimal requirement for viral budding [63, 205], is indeed sufficient to drive efficient wrapping. However, to avoid stalled partially-wrapping states and membrane rupture, the system would be confined to a relatively narrow range of adhesion strengths spanning about $2k_BT/\sigma^2$ (the + symbols in Fig. 4.6). While this result is qualitative, since the range increases in width with the particle size and the three-bead representation of the lipid molecule may lead to model membranes which are more susceptible to rupture than those comprised of a more realistic lipid, it does establish important constraints on viral evolution if budding is limited to these ingredients. However, depending on the viral system, a number of additional phenomena contribute to budding, including membrane-associated viral envelope or spike proteins, preferential budding from lipid rafts [213], the use of cytoskeletal machinery to actively drive or assist budding [187, 214] or scission [61, 188], and the ability of the virus to remodel cell membrane properties[215]. These effects can broaden the range of functional adhesion energies; e.g. using actin to drive assembly and budding [187] could enable efficient viral egress even

for adhesion energies at which spontaneous dynamics become stalled. In this case the presence of a barrier to budding could serve as a regulatory feature. The agreement between our simulation results and the elastic theory over some ranges of parameter space indicates that some or all of these effects could in principle be captured by extending existing elastic theories along the lines of Liu et. al.'s description of active endocytosis [186], but care would be required to include all relevant slow degrees of freedom near transitions between wrapping and no wrapping.

5 Simultaneous assembly and budding

5.1 Introduction

For many viruses, such as HIV, influenza or rhabdovirus [216–218], assembly and budding occur simultaneously during the last stage of the replication cycle; for others that undergo multiple budding events before exiting the infected cell, this can be the way they acquire the first lipid envelope. In general, a matrix protein, a transmembrane glycoprotein or both, mediate the interaction between the capsid proteins and the lipid bilayer. As the capsid grows by the assembly of new capsomers, the interaction with the membrane induces the bending of the bilayer around the growing capsid.

Many viruses subvert the cell machinery to induce membrane bending and final scission. Budding is actively driven in some cases by cytoskeletal activity[39, 214] or the action of the ESCRT complex [188, 219], although in other cases passive budding has been observed in modified virions lacking the ability to hijack the ESCRT complex. It has been argued that for some viruses ESCRT may accelerate vesiculation but is not essential for budding [74, 75]. In some cases, however, budding appears to occur passively and independently of these routes for wild virions, just driven by viral transmembrane proteins that are able to enhance membrane curvature [193].

It is also generally accepted that viral proteins interact with specific membrane microdomains [42] that have protein and lipid compositions different to the rest of the membrane. These microdomains are usually identified as the lipid rafts described in Sec. 1.1.3 [193, 213, 220]. Different lipid and protein concentrations with distinct intrinsic curvatures [192] and line tensions arising in the microdomain border can induce membrane bending and facilitate budding [36, 37, 221, 222].

Despite the numerous advances in virus research in the last thirty years, the properties and the kinetics of assembly and budding are still poorly understood. Only recently, new developments in microscopy, tomography and image reconstruction allowed to investigate in vivo the assembly and budding of virus particles [220, 223]. These techniques have revealed important facts related to the last stages of the viral cycle. Ivanchenko and coworkers showed that in HIV budding, the formation of multiple particles from the same site was rarely observed, and that proteins assembling into a capsid arrived directly form the cytosol or had recently attached to the membrane [220]. Findings from Briggs and coworkers also questioned the previous models of HIV assembly, suggesting that the ESCRT-mediated release

would occur concomitant and in kinetic competition with capsid assembly, so wild virions bud before the capsid is complete, being complete lattices only observed at arrested budding sites for modified virions lacking the ability to subvert the ESCRT machinery [64, 65].

Dynamic simulations might provide valuable information in this context, where there is a lack of a general and established description and experiments start to shed light into important features. Despite the great amount of work done in the field of simulation of virus assembly [161], mainly models of capsid assembly in bulk have been studied, and systems of assembly on a membrane have barely been modeled. There are however, elastic models of simultaneous assembly and budding where the free energy of the system is calculated from the membrane elastic energy and the capsomer-capsomer interaction [196]. These models can actually predict the general system behavior, such as the relation between the main magnitudes governing the process and the minimum requirements for it to happen. However, geometrical considerations and dynamical features are difficult to study with elastic theories and require the use of models with higher detail. Only very recently, a coarse grained model has been used to investigate the equilibrium states of particle assembly on a membrane [224]. We here propose a more physically realistic model to study the dynamics of the system.

In this work, we will study the assembly of icosahedral capsids on a model membrane using elastic theory for a first insight into the system behavior, and molecular dynamics simulations to get a deeper description of the geometrical and dynamical properties. The role of the genome or other viral components in the assembly process will not be taken into account, as if the virus represented here assembled prior to the genome internalization or its effect on assembly was included into the effective binding interaction between capsomers. With this model we will study capsomer assembly on a membrane domain and budding driven by capsomer-capsomer binding energy and capsomer-membrane adhesion energy. The requirements for capsomer assembly on a membrane compared to bulk assembly, the system features, and the different types of behaviors as a function of the strength of the adhesion energy and the domain line tension and size will be investigated. Finally, the dynamics of the system will also be analyzed by studying how the rate of capsomer arrival to the budding site affects the system behavior.

5.2 Methods

5.2.1 The membrane model

The membrane is modeled using the Cooke Model described in Section 3.2.2 (Eqs. 3.1 - 3.4), and the budding site is identified by the definition of an interacting domain within the membrane surface. The molecules belonging to the domain are labeled D, while those forming the rest of the membrane are referred as M. The interaction between molecules of the same type is the same, but the strength of the effective hydrophobic interaction between

molecules of different type is lower:

$$U^{ij}_{\text{hydro}}(r) = \begin{cases} -\epsilon_{ij} & ; \quad r < r_c \\ -\epsilon_{ij}\cos^2\frac{\pi(r-r_c)}{2\omega_c} & ; \quad r_c \leq r \leq r_c + \omega_c \\ 0 & ; \quad r > r_c + \omega_c \end{cases} \tag{5.1}$$

where the interaction between the molecules of the same type is given by $\epsilon_{DD} = \epsilon_{MM} = \epsilon_0$, and the cross term, ϵ_{DM}, is a parameter that controls the strength of the line tension between domains. Varying ϵ_{DM} from 0 to ϵ_0 tunes the line tension from a large value to 0.

In this way, the two domains have the same properties, such as bending coefficients or area per lipid, but a line tension arises between them because of the lower interaction between molecules of different type.

5.2.2 The virus model

Our model for capsid assembly is inspired on the model for T=1 capsids developed by Wales [180] (Sec. 3.3.1), but has been modified to allow more physically realistic interactions with the membrane. When capsomers self-assemble on a membrane, they are constrained to a two-dimensional fluctuating surface. This constraint has two important effects on the assembly dynamics, both related to the reduction of the effective degrees of freedom of the capsomers. On one hand, the interaction with the membrane involves extra terms in the Hamiltonian of the system related to the membrane dynamics, which lead to a significant alteration of the free energy landscape. On the other hand, confinement to the 2D surface restricts capsomer configurations, imposing certain angles on approaching capsomers which are determined by the local membrane curvature (Fig. 5.1).

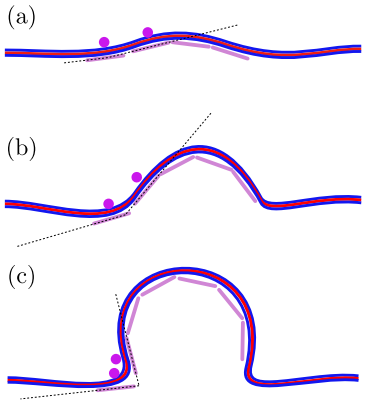

Figure 5.1: Geometry of the membrane during simultaneous assembly and budding. As budding proceeds, the angle between the growing capsid and the associating capsomers becomes more acute. In the original model the long-range repulsions between top beads do not allow sufficient orientational flexibility for capsomers to approach at such acute angles.

In models of virus assembly in bulk, the orientation specificity determines the threshold binding energy for the appearance of the kinetic trap related to malformed capsids (Sec. 3.3). High specificity increases this energy, because it implies that wrong bonds between subunits are unstable and only very strong binding energies lead to malformed capsids. However, high specificity also reduces the subunit cross section leading to lower subunit-subunit binding rates, and therefore hinders assembly [161]. We found that these constraints on assembly dynamics also apply to assembly on the membrane, but that the membrane imposes new requirements on the orientational specificity by restricting the angles of associating capsomers.

Since the capsomer has radial symmetry, the rotational degrees of freedom of a subunit can be described by two variables, a polar angle θ, for the rotation on the capsomer plane, and an azimuthal angle ϕ for rotations orthogonal to its plane (Fig. 5.2). A model for virus assembly on a membrane must have a high directional specificity on the polar angle, so that capsomers assemble with icosahedral order on the membrane surface even for low capsomer-capsomer binding energy. When adhesion between the capsomers and the membrane is strong, the model has to be specific for icosahedral geometries, otherwise, adhesion to the membrane can lead to malformed bonds that do not give rise to capsid assembly. On the other hand, since the local shape of the membrane changes considerably during budding (Fig. 5.1), the model needs to allow a wide range of approaching azimuthal angles between two capsomers, so the orientational specificity must be low for ϕ (Fig. 5.2) .

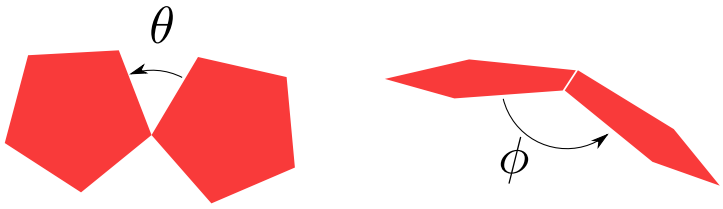

Figure 5.2: Degrees of freedom of a capsomer. The polar angle θ measures the rotation on the capsomer plane, while the azimuthal angle ϕ defines rotations orthogonal to its plane

In the following sections, the modifications of the original Wales model for assembly on a membrane are showed.

New attractor sites to increase polar specificity

In order to increase angular specificity on the polar angle, θ, new attractors are added to the edges of the capsomers (Fig. 5.3). This avoids flat hexagonal nets to equilibrate on the membrane. Edge and vertex attractors are specific, meaning that there are interactions between nearby edge attractors and nearby vertex attractors, but no vertex-edge attractor interactions. Although we did observe assembly on the membrane with the original model for carefully tuned parameters, the improved model is much more robust, meaning that it undergoes assembly over a much wider range of parameter values (Fig. 5.4).

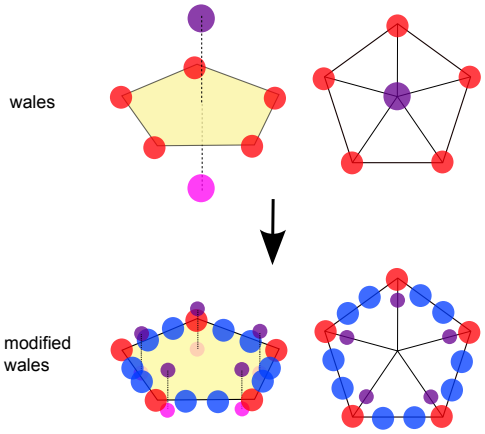

Figure 5.3: Comparison of Wales model and the modified one with the new sites. Blue circles represent the new attractive sites on the capsomer edges, and the violet and magenta circles the new repulsive sites.

The interaction potential between the attractive sites on two capsomers for the extended model is given by:

$$U_{att} = \epsilon^{v}_{\text{att}} \sum_{m} \sum_{n} \left(e^{\rho(1-R_{mn}/R_e)} - 2 \right) e^{\rho(1-R_{jk}/Re)} + \epsilon^{e}_{\text{att}} \sum_{o} \sum_{p} \left(e^{\rho(1-R_{op}/R_e)} - 2 \right) e^{\rho(1-R_{lm}/Re)} \tag{5.2}$$

where $\epsilon^{v}_{\text{att}}$ is the interaction strength between vertices sites, $\epsilon^{e}_{\text{att}}$ is the interaction strength between edge sites, R_{mn} is the distance between m and n, m running over the attractor sites on the vertices of the first capsomer , while n runs over the vertices on the second, and R_{op} is the analogous distance between the edge sites of each of the capsomers.

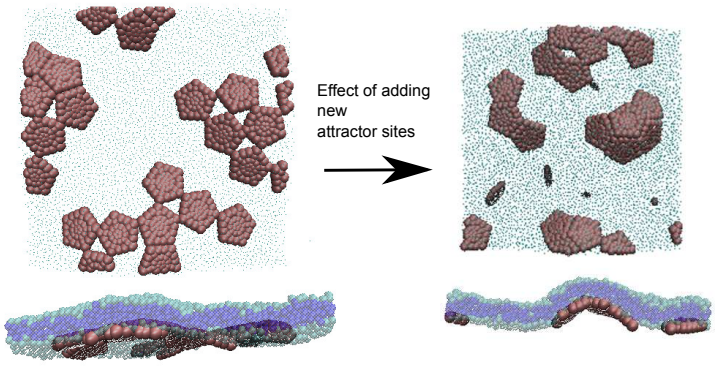

Figure 5.4: Effect of adding new attractors. With the original model capsomers tended to form defective, flat assemblages, while the modified model leads to assembly of capsids with the correct curvature and subunit-subunit interaction geometries.

New repulsive sites to reduce azimuthal specificity

Due to the changes in membrane shape caused by the thermal fluctuations or wrapping of the capsid, low specificity in the azimuthal angle ϕ is needed for two capsomers to bind while attached to the membrane. In order to increase the permitted approaching angles, the repulsive sites, which control the size of fluctuations in subunit face-face angles, are modified. The bottom and top repulsors are replaced by 5 sites at the top and 5 at the bottom closer to the capsomer plane and with a smaller interaction cutoff distance. Their projection on the plane lies on each of the pentamer radii, at a distance l to the corner, and their height, h, and l keep the same proportion as that in the original model (Fig. 5.3 and Fig. 5.5). The interaction potential between top and bottom sites on two capsomers is similar to that in the Wales model but extended to all the sites:

$$U_{rep} = \epsilon_{rep} \sum_{i=1}^{5} \sum_{j=1}^{5} \left(\frac{\sigma_t}{R_{ij}} \right)^{12} + \epsilon_{rep} \sum_{m=1}^{5} \sum_{n=1}^{10} \left(\frac{\sigma_b}{R_{mn}} \right)^{12} \tag{5.3}$$

where R_{ij} is the distance between the top sites, with i and j running over the 5 top sites of each of the capsomers, and R_{mn} the distance between m and n, m running over the bottom sites of the first capsomer and n running over the top and bottom sites of the second one. σ_t is, as in the original model, the distance between two confronted top repulsive sites in a complete capsid, and is obtained using the geometry depicted in Fig. 5.5 :

$$\sigma_t = d\sqrt{\frac{1}{10}\left(5+\sqrt{5}\right)} + 2h\sqrt{\frac{1}{10}\left(5-\sqrt{5}\right)} \tag{5.4}$$

where $d = 2l\sin(3\pi/10)$. Similarly, σ_b was initially set to the distance between the top and bottom sites of two confronted capsomers in a complete capsid, but then was adjusted to optimize assembly behavior.

This modification keeps the equilibrium angle at the same value while allowing a wider range of approaching angles (Fig. 5.6). Moreover, the reduction of the interaction cutoff considerably enhances the program efficiency.

5.2.3 Interaction between virus and membrane

The interaction potential between capsomers and membrane must account for two physical effects: adhesion and volume exclusion. Both terms need to be chosen so that adhesion plays the role of the driven force in membrane wrapping and budding, but does not reduce strongly the diffusion and mobility of the subunits on the membrane surface. This is important to enhance capsomer cross section and to allow subunits rearrangement, so kinetic traps are reduced.

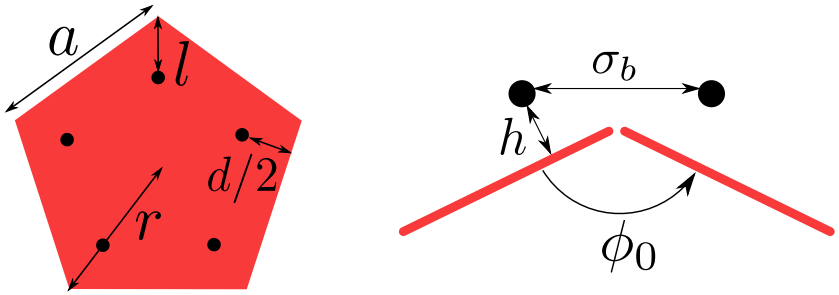

Figure 5.5: Capsomer geometry in the modified Wales model. On the left, a top view of a capsomer of radius r_v and edge length a. The projections of the new repulsive sites on the capsomer plane lie on each of the pentamer radii, at a distance l to the corner and d to the pentamer edge. Their distances from the capsomer plane, h, and l keep the same proportion as that in the original model. On the left, geometry of capsomer binding. For two adjacent pentamers on a complete capsid, the distance between to opposite repulsive sites is σ_b and

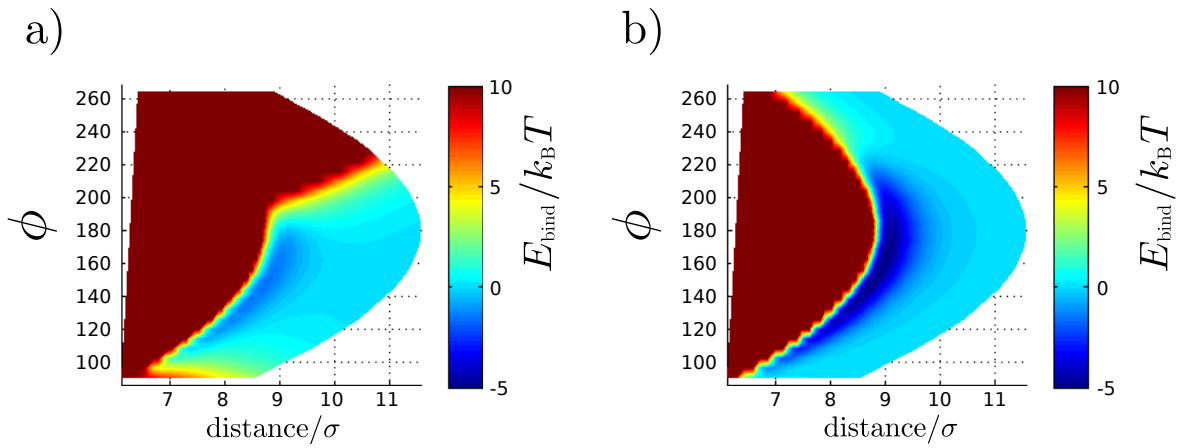

Figure 5.6: Potential energy between two capsomers associating along one edge (as in Fig. 5.2b) for different relative angles for the a) original Wales model and b) modified model. For acute angles the original model does not allow the capsomers to approach, while the modified model is much more flexible.

Adhesion interaction sites

We have in mind the Gag structure of most retroviruses to model the adhesion between the capsomers and the membrane. Gags are polyproteins with three different domains: the nucleocapsid domain that interacts with the virus genome, the capsid domain which is responsible of the protein assembly into a viral capsid, and the matrix domain that binds to

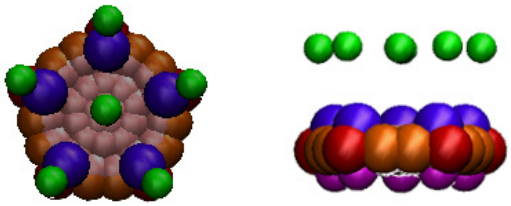

Figure 5.7: Top and side view of the capsomer. Attractive sites in red and orange, top and bottom repulsive sites in violet and magenta, capsomer excluders in pink, and membrane-capsomer interaction sites in green

the membrane via a myristate moiety that inserts into the lipid bilayer [225].

Following the Gag structure we define an interaction between the capsomer and the membrane core via six interacting points. Five of those points lie at the corners and one at the center of the capsomer and at distance L_s from its base (Fig. 5.7). These new attractors interact only with the tails of the lipid molecules of the membrane domain. The attractor-tail interaction is the same as the tail-tail interaction except that there is no repulsive component, as if the attractors were point particles with no excluded volume:

$$U_{\mathrm{ad}}(r) = \begin{cases} -\epsilon & ; \quad r < r_{\mathrm{c}} \\ -\epsilon \cos^2 \frac{\pi(r-r_{\mathrm{c}})}{2\omega_{\mathrm{c}}} & ; \quad r_{\mathrm{c}} \leq r \leq r_{\mathrm{c}} + \omega_{\mathrm{c}} \\ 0 & ; \quad r > r_{\mathrm{c}} + \omega_{\mathrm{c}} \end{cases} \tag{5.5}$$

where r is the distance between an adhesion site of a capsomer and a tail bead of a lipid.

This representation of the capsomer-membrane interaction captures the essential features of capsid protein-membrane interactions, including capsomer orientational fluctuations, while allowing for diffusion rates of capsomers adsorbed on the membrane that are high enough to enable observing assembly in computationally accessible time scales.

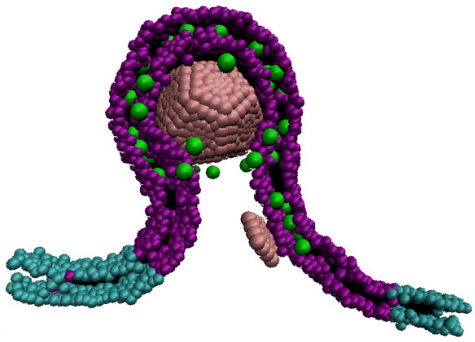

Figure 5.8: Capsomer adhesion sites (in green) distribution inside the membrane during budding.

Excluders inclusion

The repulsive and attractive sites define a bipyramidal volume exclusion between capsomers. In a model of capsomer-membrane interaction, an excluded volume between membrane units and capsomer needs to be defined to avoid overlapping. A layer of 35 beads arranged forming a pentagon is therefore added to the capsomer base (Fig. 5.7). These beads interact via an excluded volume potential with all the membrane and domain beads:

$$U_{\text{ex-M}} = U_{\text{ex-D}} = 4\epsilon_0 \left[\left(\frac{\sigma}{r-s} \right)^{12} - \left(\frac{\sigma}{r-s} \right)^{6} + \frac{1}{4} \right]. \tag{5.6}$$

with $s = (\sigma_{\text{ex}} + \sigma)/2 - 1$, and σ_{ex} as the size of the excluders.

In this way, the effective shape of the capsomer in the system is a regular pentagon of σ_{ex} thickness. This is also the representation we will use for the subunits in the figures obtained from the simulation snapshots.

5.2.4 Ghost particles inclusion to balance pressure at both sides of the membrane

The presence of capsomers on only one side of the membrane leads to a net pressure on the direction normal to the membrane. To balance this pressure, ghost spheres are introduced at both sides of the membrane. This particles experience repulsive Lennard Jones interactions with membrane head beads, but no interactions with other particles. The inclusion of these particles results in equilibration of the membrane at certain position where pressure in the normal direction reduces to zero. In this way, we can define an internal and an external volume above and below the membrane that is kept constant on average.

5.2.5 Parameters

The parameters for the membrane are chosen from the phase diagram in Fig. 3.5 (Sec. 3.2.2) [93] so that the bilayer is in a fluid state. We set the temperature of our simulations to $k_{\text{B}}T/\epsilon_m = 1.1$ and the interaction range to $\omega_{\text{c}} = 1.6\sigma_m$. The bending rigidity for these values, as showed Fig. 4.2a is $\kappa = 13.9 k_{\text{B}}T$. Values of the bilayer density can also be obtained from 4.2b.

The parameters for the virus model are set according to the phase diagrams of the original model. Following Ref. [182] and Ref. [181], we find that the optimal parameters that allow large assembly yields for a wide range of concentrations for $k_{\text{B}}T/\epsilon_v = 1$ are: $Re = 1\sigma_v$, $r_v = 5\sigma_v$, $\rho = 3$, $h = 0.1875r$, $l = 0.25r$, $\sigma_b = 0.75\sigma_t$, $L_s = 1.2r$, $\epsilon^v_{att} = 4.0\epsilon_v$, $\epsilon^e_{att} = 2.0\epsilon_v$, and $\epsilon_{rep} = 1.4\epsilon_v$.

The subindices m and v are used to refer to the units of the membrane and virus models. In order to couple both models, we use the membrane units as the system units and rewrite the parameters for the capsid assembly as a function of them. The units of energy, length,

and time in our simulations are respectively $\epsilon_0 = \epsilon_m$, $\sigma = \sigma_m$ and τ_0. The values of the capsid parameters are chosen so that the total energy of assembly exceeds the bending energy of wrapping the capsid (see Sec. 5.3). In this way, the unit of energy of the assembly model is set to $\epsilon_v = 5\epsilon_m = 5\epsilon_0$, the length and time parameters are the same of those of the membrane model, $\sigma_v = \sigma$ and $\tau_v = \tau_0$. Since the units of ϵ_0 are $[m_0\sigma^2/\tau^2]$, the energy scale implies that $m_v = 5m_0$. m_v is the mass of each interaction site of the capsomer, since there are a total of 66 sites, $m_{\text{capsomer}} = 66m_v = 330m_0$. The effective radius of the capsid can be estimated from the average penetration of the spike into the membrane plus the radius of a sphere inscribed in a dodecahedron, $r_i = \frac{a}{2}\sqrt{\frac{5}{2} + \frac{11}{10}\sqrt{5}}$ where a is the edge of the pentamer and is given by Eq. 3.7. This results with $R = r_i + L_s/2 \approx 1.91r_v$.

The remaining parameters can be assigned physical values by setting the system to room temperature, $T = 300K$, and noting that the typical width of a lipid bilayer is around 5 nm, and the mass of a typical phospholipid is about 660 g/mol. The units of our system can then be assigned as follows: $\sigma = 0.9$ nm, $m_0 = 220$ g/mol, $\epsilon_0 = 3.77 \times 10^{-21}\text{J} = 227\text{g}\mathring{\text{A}}^2/\text{ps}^2\text{mol}$, and $\tau_0 = \sigma\sqrt{m_0/\epsilon} = 8.86$ ps. The effective radius of the capsid in the system units is $R = 9.5\sigma$.

5.3 Energetics of assembly on a membrane

Virus assembly during budding involves the binding of capsomers to the membrane at the same time that they assemble to form a new capsid. These two processes lead to the bending of the surface while it wraps the growing capsid. The balance between two energy scales will determine the feasibility of the process: the energy gain on assembly has to balance the energy loss needed to bend the membrane. Both processes are coupled by the capsomer adhesion to the membrane. Even if the gain from assembly balances the bending cost, if the adhesion strength is low, the energy gain by attaching to the membrane will not be enough to compensate the loss upon bending, and the capsid will assemble in the bulk. On the other hand, too high adhesion energies might give place to systems where the gain upon assembly is negligible compared to the attaching to the membrane, and capsid formation would not be promoted. It is thus important that not only the energy of capsid formation exceeds the bending energy but also that the coupling term, the adhesion strength, takes the appropriate values.

As commented above, the units of the system are set to those of the membrane, but in order to relate them with the ones from the capsid assembly model, both energy scales need to be compared. In this section, we will estimate the different energy scales involved in the assembly and budding process with a twofold aim: first, the analysis of the energetics will give us a first approach to the system behavior, and secondly, this will allow us to set the appropriate range of parameters in our model.

In order to get budding, the values of the parameters of the capsid model are chosen from the comparison between the bending energy of the membrane during wrapping, and that of assembly. The bending energy of a spherical vesicle is $E_{\text{bend}} = 8\pi\kappa$. Here we are considering only the energy after budding, the extra term from the rim area around the partially wrapped capsid is not included, so only the equilibrium state is studied. For the set of parameters we use, the bending coefficient of the membrane is $\kappa \approx 13.9 k_B T$, so the bending energy per capsomer is $e^{\text{bud}}_{\text{bend}} = E^{\text{bud}}_{\text{bend}}/12 \approx 29\epsilon_0$.

On the other hand, the energy of assembly is calculated from the binding energy per bond between capsomers F_{bind}:

$$F_{\text{bind}} = -k_B T \log\left[1 + \int_\sigma^\infty dr \left(e^{-\mathcal{U}^i_{\text{att}}(r)} - 1\right)\right] - TS_{\text{bind}}. \tag{5.7}$$

where U^i_{att} has been defined in Eq. 5.2 and i can refer to the vertices or to the edges contribution. Although the interaction between two attractors is usually mediated by the presence of the other attractors that are not directly associating, in our estimation for the binding energy we neglect that part, so we will obtain a lower value for the energy. There is also a entropic part, S_{bind}, related with the loss of rotational entropy upon assembly, that we will consider as an extra term. Within these assumptions Eq. 5.7 is solved for both contributions for the parameters set chosen above. The results as a function of ϵ_v are shown in Fig. 5.9. For the range of ϵ_v we are interested, we can approximate the free energy to a linear function, so we obtain $F^v_{\text{bind}} \approx 1.42 - 3.81\epsilon_v$ for the vertices, and $F^e_{\text{bind}} \approx 0.92 - 1.80\epsilon_v$ for the attractors on the edge. The total binding energy per capsomer can be written as $e_{\text{bind}} = 5/2F^v_{\text{bind}} + 10/2F^e_{\text{bind}} \approx 8.2 - 18.5\epsilon_v$

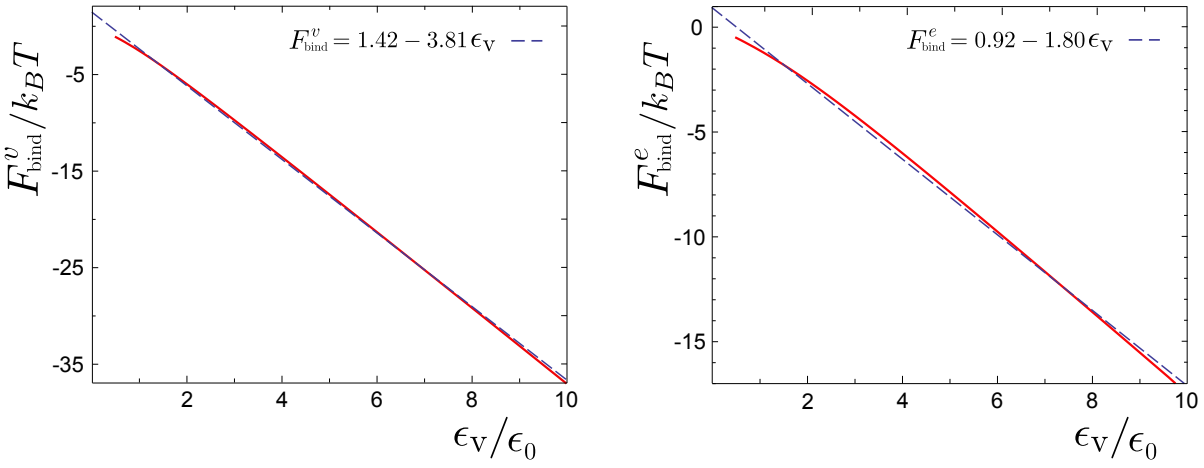

Figure 5.9: Assembly energy estimation as a function of the attractors interaction strength obtained from Eq. 5.7 for vertices (on the left) and edges (on the right) interaction. The results can be approximated in both cases by a linear function for the range of parameters relevant for the system.

For assembly and budding to occur, the gain in binding free energy must be greater than the loss upon bending, so $|e_{\text{bind}} - TS_{bind}/n_{\text{capsomer}}| > |e_{\text{bend}}|$ must hold true. But also, as we commented above, the coupling term, which is the adhesion energy, must be properly set.

The adhesion energy is estimated from the calculation of the mean interaction between a spike and the tail beads inside its interaction range. The number of beads depends on the spike penetration, so the energy is averaged for the whole penetration range (Fig. 5.10).

$$e_{\text{ad}} = \frac{6}{\Delta p} \int_{p_{\min}}^{p_{\max}} E_{\text{ad}}(p) \mathrm{d}p. \tag{5.8}$$

where $E_{\text{ad}}(p)$ is the adhesion energy for a given penetration and is obtained from the sum over all the interactions between the spike and every tail bead lying inside the interaction volume defined by the cutoff radius (Fig. 5.10) :

$$E_{\text{ad}}(p) = 2\pi\rho \int_{z_{\min}(p)}^{z_{\max}(p)} \mathrm{d}z \int_{0}^{R_{\max}(z)} U(r) \mathrm{d}R \tag{5.9}$$

where ρ is the density of lipids in the membrane and the factor 2π comes from the rotational symmetry in the plane of the membrane. The factor 6 in Eq. 5.8 comes from the number of spikes per capsomer. The density of tail beads is obtained from the surface density of lipids, η (Fig. 4.2), $\rho = 4\eta/h_0$, where 4 is the number of tail beads from two amphiphiles confronted in the lipid bilayer and h_0 is the width of the hydrophobic part of the layer formed by the tails.

For the actual parameters, $e_{\text{ad}} \approx -71.6\epsilon_{\text{ad}}$. For values of ϵ_{ad} between $0.1\epsilon_0$ and $1.0\epsilon_0$, e_{ad} ranges from -14.3ϵ_0 to -71.6ϵ_0. We summarize the estimations for the energies in Fig. 5.11a, and we observe that all of them are of the same order. If the energy gain during assembly compensates the energy loss upon bending, assembly and budding can occur depending on the adhesion strength: for low values of ϵ_{ad} the capsomers would assemble in bulk, while for large values, adhesion will be favored over assembly, and capsids will not be formed. For intermediate values we would expect assembly of well-formed capsids and budding. The relation between these values and their effect on the system behavior has been summarized in Fig. 5.11b , and those assumptions have been used to obtain a phase diagram of the system behavior as a function of the energy estimators (Fig. 5.11c).

Three main behaviors are observed and a range of parameters for right assembly and budding is obtained. Following Fig. 5.11c, we set $\epsilon^w = 5\epsilon_0$. Although this value predicts assembly and budding, we only expect it to happen in a narrow range. However, our system will also be affected by the effect of the domain. The energy stored in the domain e_{domain} can induce budding, compensating the bending energy, so budding is enhanced even for low adhesion strengths. The value of e_{domain} will shift the range of energies where budding is accessible, therefore, it is important to get a good estimator also for this energy. When a domain is defined in the membrane, a line tension arises between the two different regions

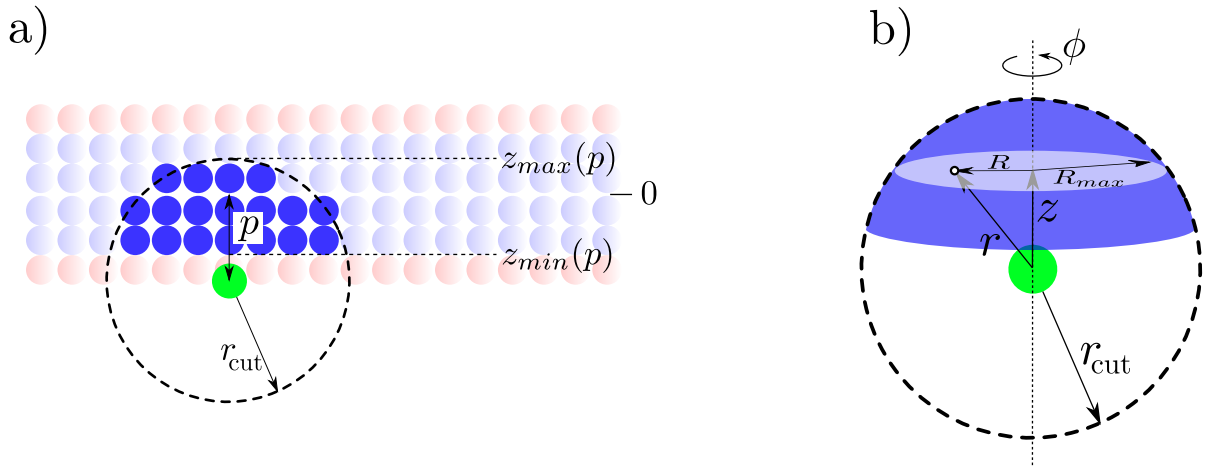

Figure 5.10: Geometry used for the calculation of the adhesion energy. a) The schematic shows a slice of a membrane with a spike in green at a given penetration p. The lipid tails are represented in blue and the heads in red. The tail beads that lie inside the spike interaction volume V_{int} are showed in solid blue and they are confined between z_{max} and z_{min} in the z-direction. The potential cutoff is given by $r_{\text{cut}} = r_c + \omega_c$. b) Geometry used for the integration of the adhesion energy . The energy contribution of every point inside a disk of radius $R_{\text{max}}(z)$ and width dz is integrated inside the interaction volume (represented in blue).

proportional to the length of the interface:

$$e_{\text{domain}} = 2\pi r_{\text{domain}}\tau \tag{5.10}$$

where r_{domain} is the radius of the interacting domain and τ is the line tension per unit of length. This energy will be reduced if the length of the interface between the two domains decreases. This can only be achieved if the domain protrudes out of the plane of the membrane, with e_{domain} being 0 in the limit where the whole domain buds and separates from the rest of the membrane. However, the bending energy of forming a vesicle competes with the line tension for the domain to bud out of the membrane. The balance of both energies will determine the feasibility of the process and the final steady state.

The energy stored in the edge can be estimated from the difference between the stress components normal and tangent to the interface [201]. We calculate it using the coordinate system shown in Fig. 5.12, where z is normal to the membrane and y is normal to the interface at a given point and equals 0 at the interface. We integrate the transversal and normal components of the stress as a function of y for the whole membrane height (Fig. 5.12):

$$\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_x(h) - \sigma_y(h))\mathrm{d}x\mathrm{d}h = \int_{-h_0/2}^{h_0/2}\int_{-r_{\text{cut}}}^{r_{\text{cut}}}(\sigma_x(h) - \sigma_y(h))\mathrm{d}x\mathrm{d}h \tag{5.11}$$

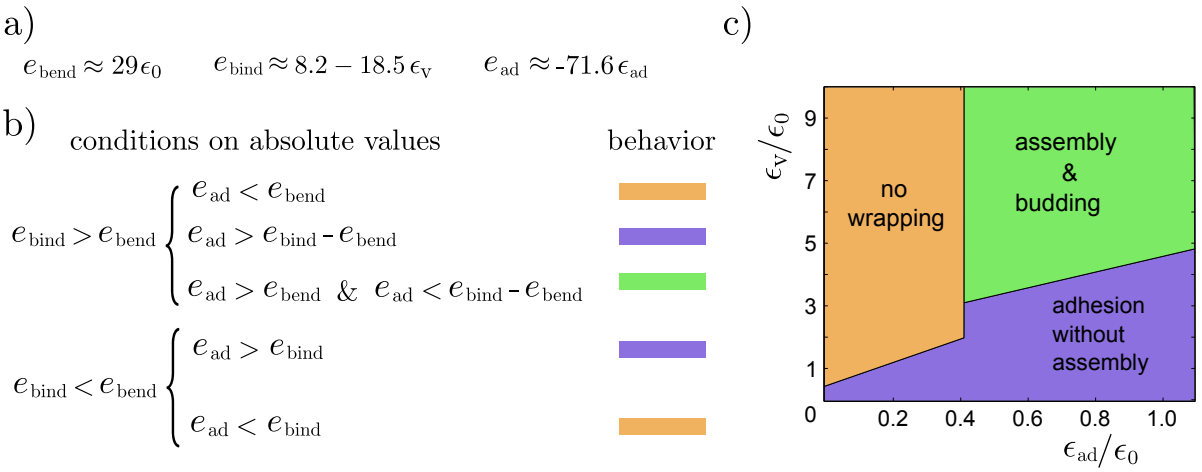

Figure 5.11: Energetics of assembly and budding. a) Summary of the values of the system energies. b) Conditions for the final behavior of the system in equilibrium where the energy estimators refer to their absolute values. c) Phase diagram of the system behavior as a function of the binding and the adhesion strength estimators. In orange the region where the membrane does not totally wrap the assembled capsid. In violet, the region where the adhesion energy is more favorable than the binding, so the capsomers just attach to the membrane instead of assembly. In green, the intermediate region where the energies are balanced so capsid assembly and budding take place.

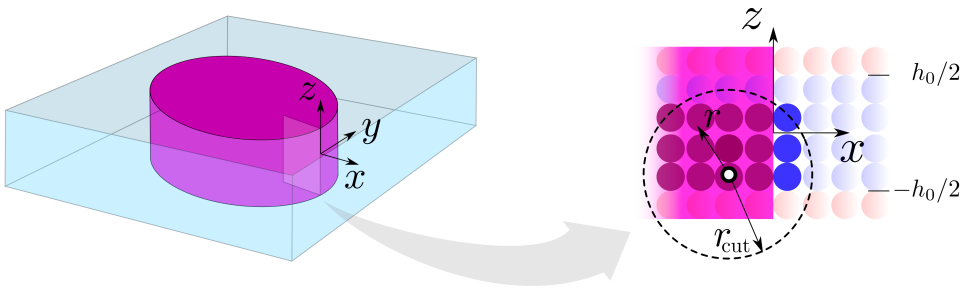

Figure 5.12: Geometry for the calculation of the line tension. On the left, a circular domain (violet) lies inside a membrane (blue) on the x-y plane. On the x-z plane, where tension is calculated, the domain border is normal to x and tangent to y. On the right, a schematic of a slice of the domain-membrane boundary is shown, where the integration volume for a given point inside that boundary is depicted. The tension is integrated for the whole boundary for $x < |r_{\text{cut}}|$ and $z < |h_0/2|$ where the membrane midplane is at $z = 0$ and the membrane width is h_0

where σ_x and σ_y are the components of the diagonal microscopic stress tensor in the x and y directions and h_0 is the membrane width. The integral limits in x are rewritten as a function of the interaction range, for $x > |r_{\text{cut}}|$, $\sigma_x = \sigma_y$, so the contribution to the integral outside this range is 0. The stress components in a pairwise potential [136] and for the geometry from

Fig. 5.10 are defined as :

$$\sigma_x(h) = \frac{1}{\Delta V(h)} \sum r_x^{ij} f_x^{ij} = \frac{\rho}{V_{\text{int}}(z)} \int_{z_{\min}(h)}^{z_{\max}(h)} \mathrm{d}z \int_0^{2\pi} \mathrm{d}\phi \int_0^{R_{\max}(z)} r_x f_x(r) \mathrm{d}R \tag{5.12}$$

$$\sigma_y(h) = \frac{1}{\Delta V(h)} \sum r_y^{ij} f_y^{ij} = \frac{\rho}{V_{\text{int}}(z)} \int_{z_{\min}(h)}^{z_{\max}(h)} \mathrm{d}z \int_0^{2\pi} \mathrm{d}\phi \int_0^{R_{\max}(z)} r_y f_y(r) \mathrm{d}R \tag{5.13}$$

where $f_x(r) = -\nabla_x U^{ij}_{\text{hydro}}$ and $f_y(r) - \nabla_y U^{ij}_{\text{hydro}}$ are the forces on x and y upon the point where the stress is calculated due to a particle at distance r. As we defined in Sec. 5.2.1, the interaction between particles belonging to different domains has a different strength that the one between like particles. U^{ij} will take values of U^{DD}, U^{MM} or U^{DM} depending on the point where the stress is calculated and that where the contribution to the total force comes from. Finally, $V_{\text{int}}(h)$ is the volume of the layer formed by the tail beads that lie inside the interaction range and depends on the z-coordinate of that point:

$$V_{\text{int}}(h) = 2\pi \int_{z_{\min}(h)}^{z_{\max}(h)} \mathrm{d}z \int_0^{R_{\max}(z)} \mathrm{d}R \tag{5.14}$$

When solving numerically Eq. 5.10 with Eqs. 5.11 - 5.14, we obtain that the line tension is a linear function of the strength between the particles of different domains ϵ_{DM} and the domain radius, r_{domain} (Fig. 5.13):

$$e_{\text{domain}} = 2\pi r_{\text{domain}} \tau_0 (\epsilon_{\text{DM}} - 1) \tag{5.15}$$

with $\tau_0 = 7.408\epsilon_0/\sigma$ for the parameters defined in Sec. 5.2.5.

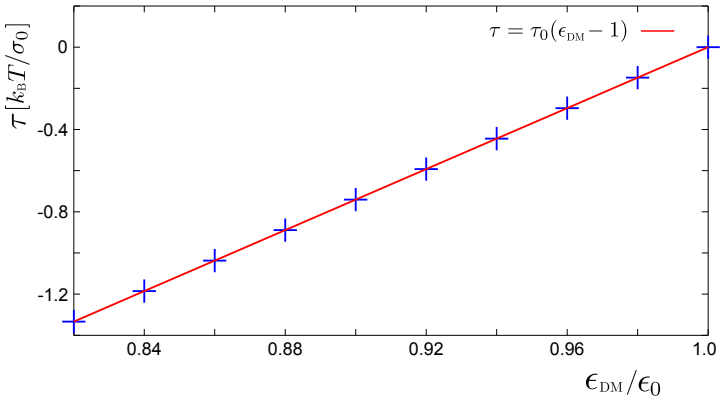

Figure 5.13: Values for the line tension estimation as a function of the interaction strength between particles belonging to different domains obtained from Eq. 5.10. The results are approximated to a linear function.

With $\epsilon_v = 5\epsilon_0$ we obtain a phase diagram as a function of the line tension and the adhesion strength using the same conditions from Fig. 5.11 but taking into account the effect of the line tension into the bending energy, $e_{\text{bend}} = e_{\text{bend}} + e_{\text{domain}}$, where e_{domain} takes negative values in the region under study.

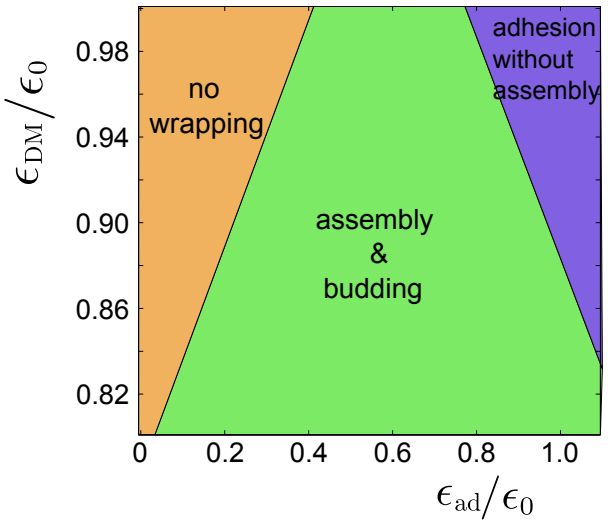

Figure 5.14: Energetics of assembly and budding on a domain taking into account the line tension. Phase diagram of the system behavior as a function of the line tension (proportional to ϵ_{DM}) and the adhesion strength for the conditions summarized in Fig. 5.11b but with the energy of the line tension included in the bending energy term, $e_{\mathrm{bend}} = e_{\mathrm{bend}} + e_{\mathrm{domain}}$. In orange the region where the membrane does not wrap the assembled capsid. In violet, the region where the adhesion energy is more favorable than the binding, so the capsomers just attach to the membrane instead of assembly. In green, the intermediate region where the energies are balanced so capsid assembly and budding take place.

In the following study we will set the bending energy and the attractor energy strength constant at the already chosen values that predict assembly on the membrane, $\epsilon_v = 5\epsilon_0$ and the parameters from Sec. 5.2.5. We will study the process as a function of the adhesion energy, the strength of the interaction between membrane and domain tails, and the radius of the domain.

5.4 Simulations

Molecular dynamics simulations were performed on GPUs using HOOMD-blue package version 0.10.1 [129]. The original algorithm used to simulate the membrane in the previous chapter was replaced in order to adapt it to the integrators available in HOOMD.

The membrane was simulated at constant temperature and lateral tension using a two steps Verlet algorithm (Sec. 2.3.1) with a Nosé-Hoover thermostat [226] (Sec. 2.3.3) to maintain constant temperature, and an Anderson barostat [119] (Sec. 2.3.3) to simulate a tensionless infinite membrane. The Anderson barostat implemented in HOOMD keeps the pressure constant, so it was modified to simulate at constant tension. The z-axis in our case, is free to fluctuate and does not contribute to the pressure. While the barostat adapts the velocities and positions of every particle according to the changes in the volume, the modified

barostat rescales only their x and y components. This statistical ensemble reproduces the conditions of the cellular environment, where temperature is constant around 37 °C and the membrane is tensionless on average. The time step was $\Delta t = 0.01\tau_0$, the membrane is coupled to the thermostat and barostat via their characteristic times, $\tau_T = \sqrt{Q/gk_BT} = 0.4\tau_0$ and $\tau_P = \sqrt{W/gk_BT} = 0.5\tau_0$ respectively. The reference pressure, P_0, was set to 0, to simulate a tensionless membrane.

Each capsomer was simulated as a rigid body using Langevin Dynamics [115] at constant volume and temperature (Sec. 2.3.3). The algorithm uses quaternions notation for the rigid body dynamics (Sec. 2.3.2) and was modified specifically for this work. The friction and random forces are applied to the center of mass of each rigid body and the damping coefficient was approximated to that of a spherical particle, so it is the same for the three main axis of the capsomer. This modification gives a closer approximation to the rotational motions of globular proteins. The changes in the box size generated by the barostat to ensure constant lateral tension in the membrane were coupled to the dynamics of the rigid body at each step, so the centers of mass of the rigid bodies were adapted accordingly.

As stated above, the units of energy, length, and time in our simulations are respectively $\epsilon_0 = \epsilon_m$, $\sigma = \sigma_m$ and τ_0, and the parameters of the model for capsid assembly are written as a function of them. The Langevin dynamics applied to the capsomers are defined for $k_BT_v = \epsilon_v = 5\epsilon_0$ and the translational and rotational damping coefficients are set to $\gamma = 0.02m_0/\tau_0$ and $\gamma_{\text{rot}} = 0.04m_0\sigma^2/\tau_0$ respectively.

In order to simulate an infinite membrane, periodic boundary conditions were employed for the lateral dimensions and a wall was placed on the bottom of the box so the capsomers remain inside the compartment defined by the membrane unless they bud through it.

The membrane was comprised of $n = 16,200$ lipids divided into two different domains. The dependence on the system behavior as a function of the interacting domain size is studied with the simulations. For every domain size, an initial bilayer configuration was relaxed by MD and then placed normal to the z-axis in a cubic box of side-length $L_x = L_y = 90\sigma$ and $L_z = 100\sigma$.

The capsomers are introduced in the box in two different ways. The initial configuration has always three capsomers placed at some distance below the membrane. In the simulations used to study the equilibrium states, a new capsomer is injected once the previous subunits are assembled into a single capsid. When the effect of the monomer concentration on the membrane and in the surrounding cytosol is analyzed, the capsomers are injected at a constant frequency. Simulations for different values of that frequency are performed to mimic the effect of different concentrations. For injection rates above the average assembly rate, the concentration of free capsomers increases with this frequency. At the limit of infinite frequency, all the capsomers are injected at a time.

For each set of parameters, the results from four independent simulations are averaged to obtain the mean behavior of the system.

5.5 Results

The previous sections give a simplified prediction for the effect of line tension in the other parameters involved in assembly and budding, but the simulations will give a more precise description, since they include effects not accounted for in the simplified prediction and allow to get an insight into the dynamics. The capsid assembly has been studied in the presence of a microdomain and without it, and the behavior of the system has been analyzed as a function of the line tension and the monomer injection rate for different adhesion strengths.

5.5.1 Capsomers assembly on the membrane surface

When simulating the system without the domain, the capsid does not usually assembly on the membrane beyond half its size, in our case with T=1, this means that only six particles assemble. For low adhesion strengths, the seventh particle usually assembles but it detaches from the membrane. Sometimes the capsid can even grow further, but the membrane does not wrap it (Fig. 5.15). At high adhesion strengths however, the capsomer binds to the capsid but it does not bend the membrane, and remains attached but without closing the icosahedral structure (Fig. 5.16). At intermediate adhesions it does not even stay attached for long times.

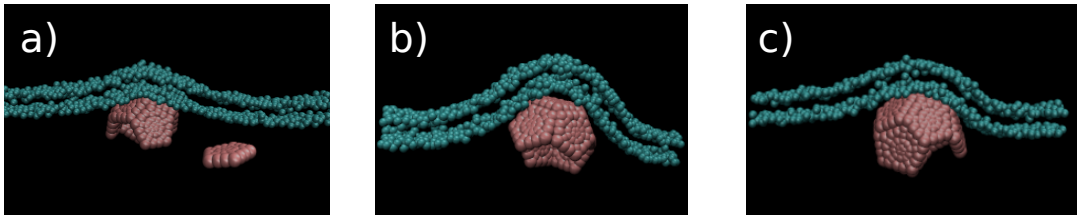

Figure 5.15: Assembly of the seventh particle for low adhesion strengths. a) capsid of 6 capsomers and seventh free pentamer, b) capsid of seven capsomers, the membrane does not wrap the whole capsid. c) capsid of nine capsomers, when assembling into the capsid, some pentamers detach from the membrane.

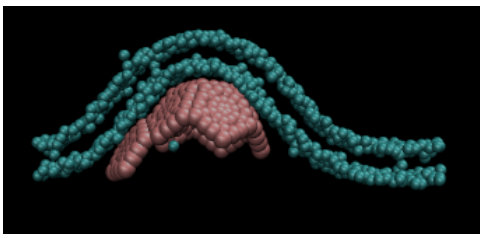

Figure 5.16: Assembly of the seventh particle for high adhesion strengths. The particle assemble but does not bend the membrane and therefore, cannot adopt the right position into the icosahedron.

The capsid cannot grow beyond half its final size due to the geometry of the system at that point. As shown in Fig. 5.17, the conformational changes that the membrane undergoes in order to wrap the particle beyond the half size are greater and are related with a higher increase in the bending energy than in previous steps.

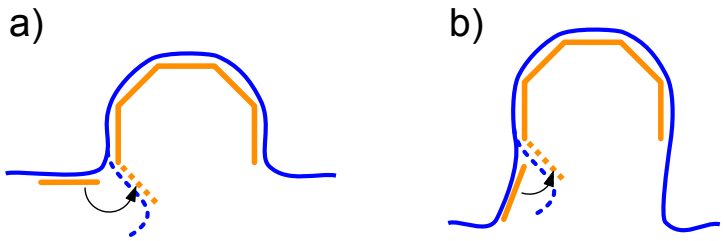

Figure 5.17: Membrane conformational change required for capsomer assembly when a) short necks and strong kinks and b) long necks and soft kinks.

Proper assembly of a capsomer beyond the half capsid on a membrane without domains has only been observed to happen via the coordinate action of two or more capsomers, as free monomers, dimers or bigger aggregate. The simultaneous attachment can provide the energy to promote the membrane conformational changes needed to wrap the capsid beyond the half state, which usually comprises high curvatures (Fig. 5.18). In order to exploit this advantage, simulations with many free capsomers have been studied. However, even when an aggregate of 8 or more capsomers is found, the membrane does not wrap the particle entirely and budding does not occur. When assembly proceeds from many free pentamers, the system is trapped into monomer starvation, two partial capsids usually equilibrate, and the membrane does not wrap them (fig. 5.19).

Membrane configurations with long necks and low curvatures would reduce the curvature changes that the seventh and next particles need to induce on the membrane when assembling. These configurations can be easily obtained using elements that induce curvature: a domain is a suitable candidate, but other possibilities could involve proteins associated to the capsid that enhance bending [158], or models of virus assembly with curved capsomers, or with a higher number of proteins forming the capsid, so the changes on the membrane that a single particle needs to induce are smaller.

In the following sections a domain is defined within the membrane as platform for capsid assembly and budding.

5.5.2 Effects of the domain on membrane dynamics

Before studying the capsid assembly on a domain, we study the effects of the domain alone on the membrane dynamics. As we commented in Sec. 5.3, in the presence of a domain the interfacial energy decreases when the length of the domain interface is reduced. Since the membrane is not confined in the normal direction, the border length can be reduced if the

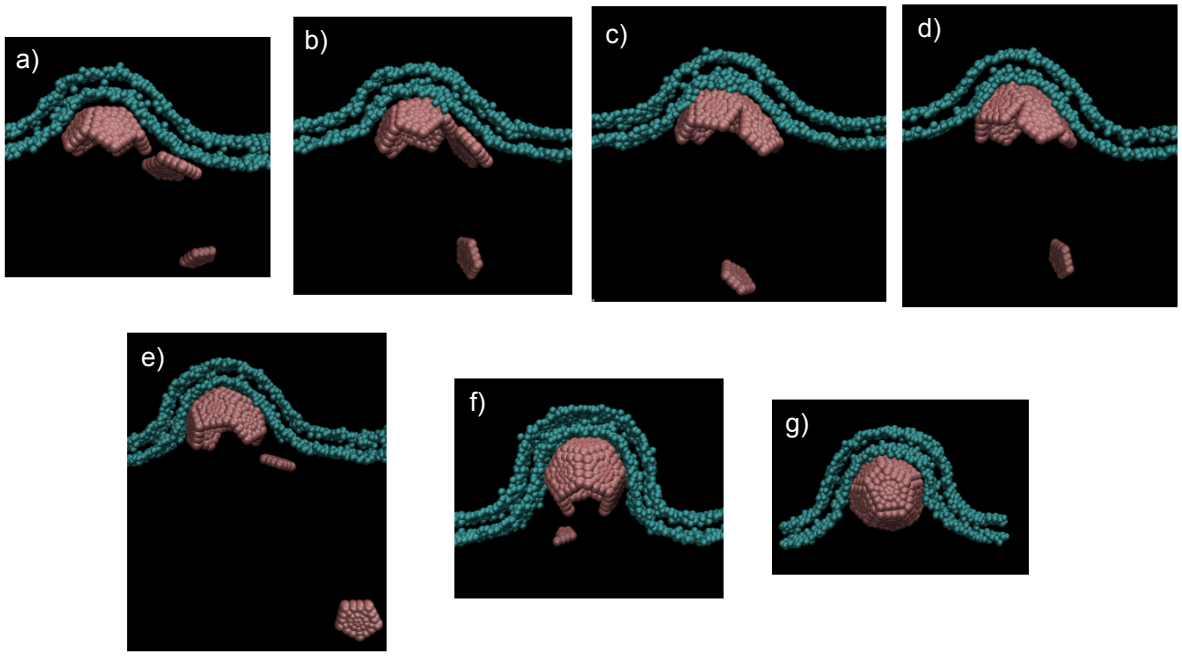

Figure 5.18: Assembly of the seventh particle via dimer assembly. a) dimer approaches a capsid formed by 6 capsomers. b),c),d) dimer assembly, e),f),g)assembly of the whole capsid but without membrane wrapping.

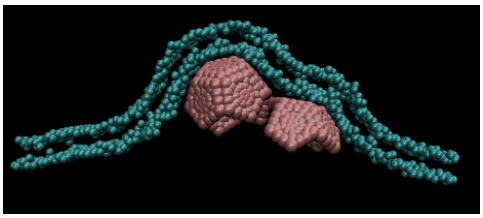

Figure 5.19: Assembly of 12 free pentamers on the membrane forming two stable partial capsids

domain protrudes out of the plane with the corresponding penalty on bending energy. For high enough tensions, related with a significant difference between interparticle, ϵ_{DM}, and like-particles, ϵ_0, interaction strengths, thermal fluctuations alone can induce curvature of the domain through small protrusions, and a vesicle can form and bud (Fig. 5.20). When ϵ_{DM} and ϵ_0 are similar however, the domain segregation cannot compete with the entropy penalty of mixing, so the particles forming the domain dissolve in the membrane, and no domain is longer defined.

The energy stored in the domain is also proportional to the domain size (Eq. 5.10), the larger the domain, the higher the energy related to the line tension. However, the bending energy of a vesicle does not depend on its radius, so larger domains present spontaneous budding for smaller values of the line tension (Fig. 5.21a).

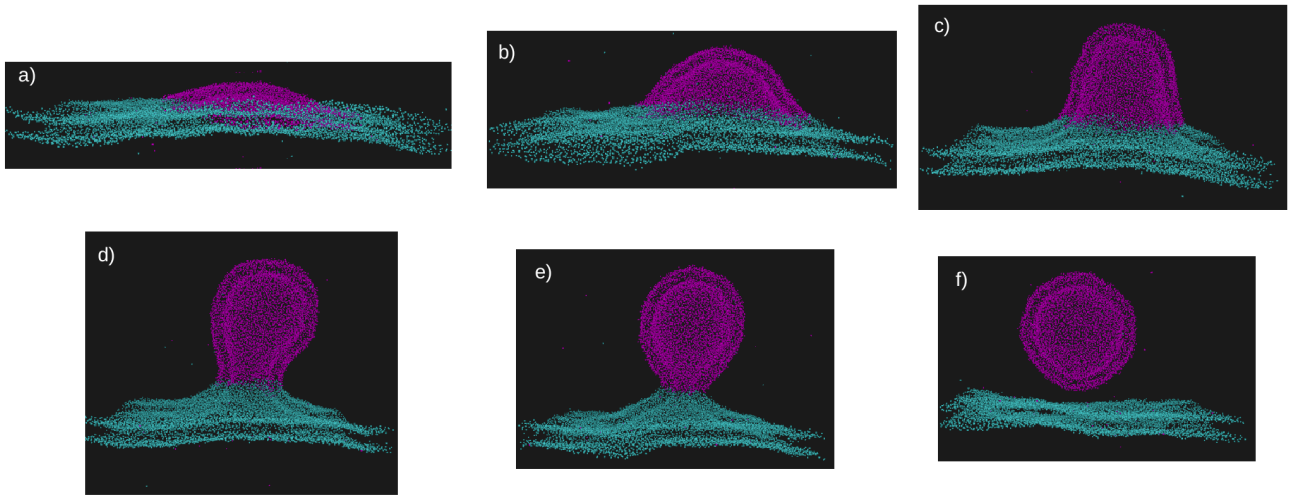

Figure 5.20: Spontaneous budding of the whole domain for $r_{\text{domain}} = 35\sigma$ and $\epsilon_{\text{DM}} = 0.8\epsilon_0$.

The spontaneous budding as a function of the the domain size and the line tension is shown in the phase diagrams in Fig. 5.21 for the energy estimations and for the simulation results. In the simulations, thermal fluctuations facilitate domain budding for smaller line tensions than those predicted analytically, because they produce membrane deformations of the length of the domain that provide an extra energy for membrane bending.

As we can see in the phase diagram, domains in equilibrium with the surrounding membrane are obtained for a specific range of the interparticle interaction strength that depends on the domain size. For higher values of ϵ_{DM} the domain dissolves, and for lower values, it buds out of the membrane.

5.5.3 Capsomer assembly on a domain

To study assembly on a domain, we set the parameters using Fig.5.21 so that the interfacial energy is high while the domain is in equilibrium within the membrane at the temperature of the simulation, $\epsilon_{\text{DM}} = 0.85\epsilon_0$ and $r_{\text{domain}} = 35\sigma$. Depending on the adhesion strength, different behaviors are observed (Fig. 5.22). Low adhesions promote long and soft necks that reduce the domain border until it is not stable stable anymore and the whole domain buds before the capsid is complete (Fig. 5.23). For intermediate strengths, necks are less elongated, and the membrane adopts shapes with areas of higher curvature around the capsid while increases the contact area, the particles assemble one after the other in a whole capsid . Once the capsid is formed it remains stable for long times (longer for high adhesions strengths), until a fluctuation fuses the neck and the capsid buds surrounded by part of the domain, leaving the rest of the domain on the membrane (Fig. 5.24). High values of epsilon promote short necks and acute kinks on the membrane rim around the growing capsid. These

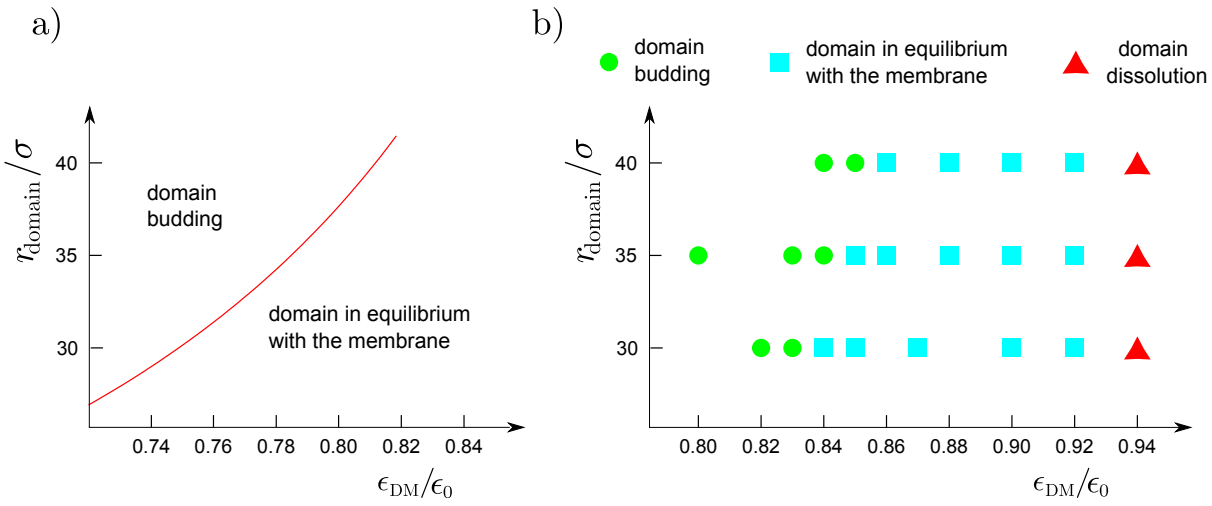

Figure 5.21: Conditions for spontaneous budding of the whole domain. a) Transition between spontaneous domain budding and domain at equilibrium with the membrane for $e_{\mathrm{domain}} = e_{\mathrm{bend}}$ as a function of the domain radius and the line tension (proportional to ϵ_{DM}. b) Phase diagram of the domain behavior as a function of the domain radius and the line tension (proportional to ϵ_{DM} from simulations at $k_{\mathrm{B}}T = 1.1\epsilon_0$. The transition between the events representing domain budding (green circles) and domain equilibrium (light blue squares) is shifted towards higher values of ϵ_{DM} compared to the analytical line. It should be notice that the effect of the thermal fluctuations is only affecting the simulation results. Thermal fluctuations can take the domain out of the equilibrium and trigger budding for smaller line tensions than those predicted by the analytical model.

configurations show very high curvatures so that the last capsomers either do not cross the kinks or cannot bend the membrane to the position where they fit into the capsid. Capsids with more than 10 or 11 capsomers are not observed, and budding does not occur (Fig. 5.25).

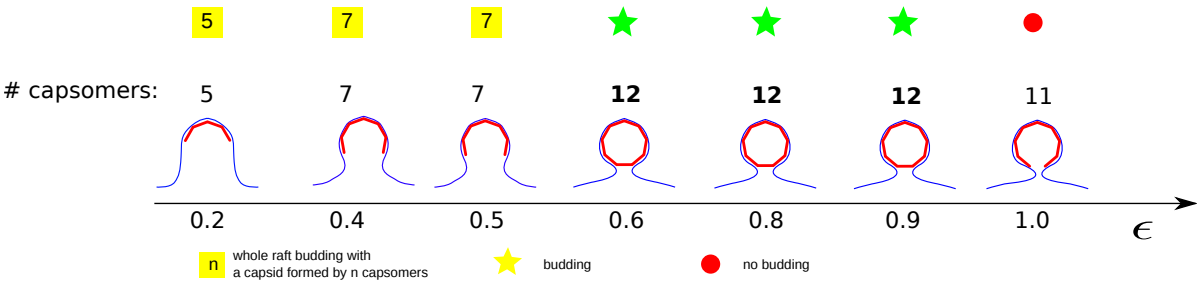

Figure 5.22: Budding behavior. Low adhesion strengths promote budding of the whole raft and high adhesion strength prevent the particle for budding. For intermediate adhesions the capsid assembly correctly and buds.

Sometimes during assembly a hole in the capsid is formed, since the hole is surrounded by pentamers, the membrane can wrap it and the final capsid buds with a hole on it (Fig.

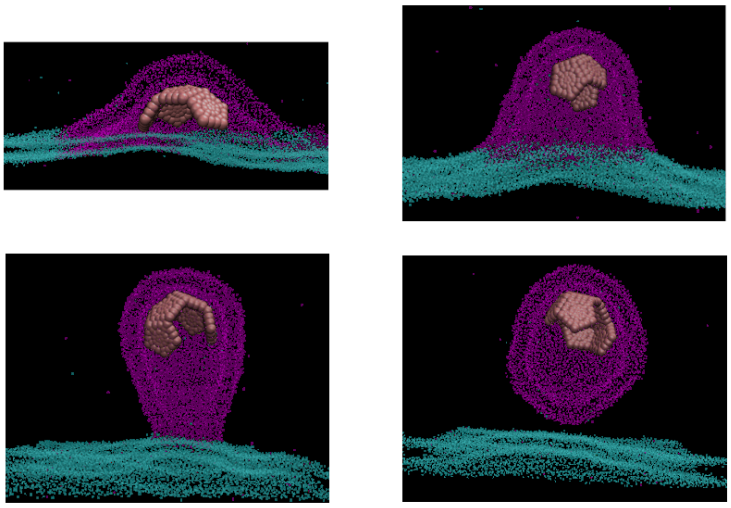

Figure 5.23: Budding of the whole raft containing a partial capsid.

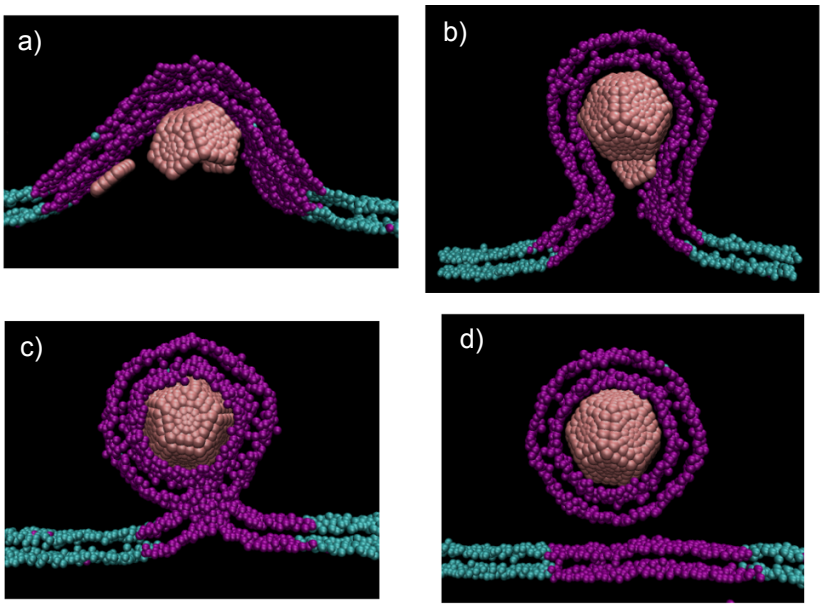

Figure 5.24: Capsid assembly and budding.

5.26). In other cases the interaction with the membrane is such that the neck is too narrow for the last particle to enter, and the capsid also buds with a hole.

The effect of the thermal fluctuations can be compared to that of the domain curvature generated by the adhesion of a growing capsid, but they differ in an important feature: the length of the deformations. The largest thermal fluctuations are of the size of the simulation box, and those driving domain budding have a wavelength of the order of the domain diameter, specifically those with $\lambda_{k_BT}/2 = 2r_{\text{domain}}$ will produce the greatest undulations of the

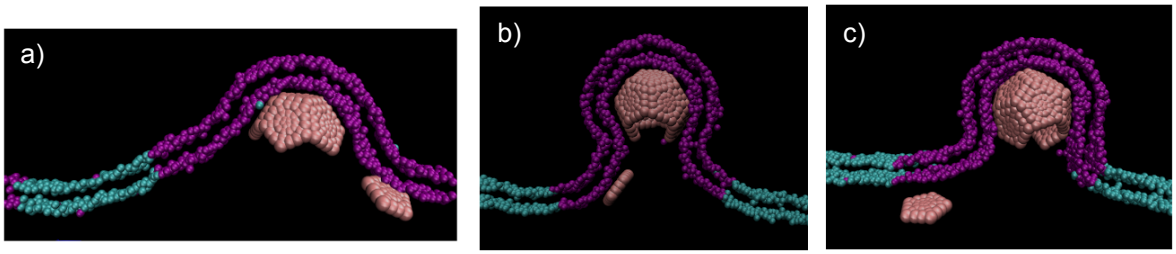

Figure 5.25: Arrested state. A capsid formed by 10 pentamers cannot grow further because new capsomers cannot assemble.

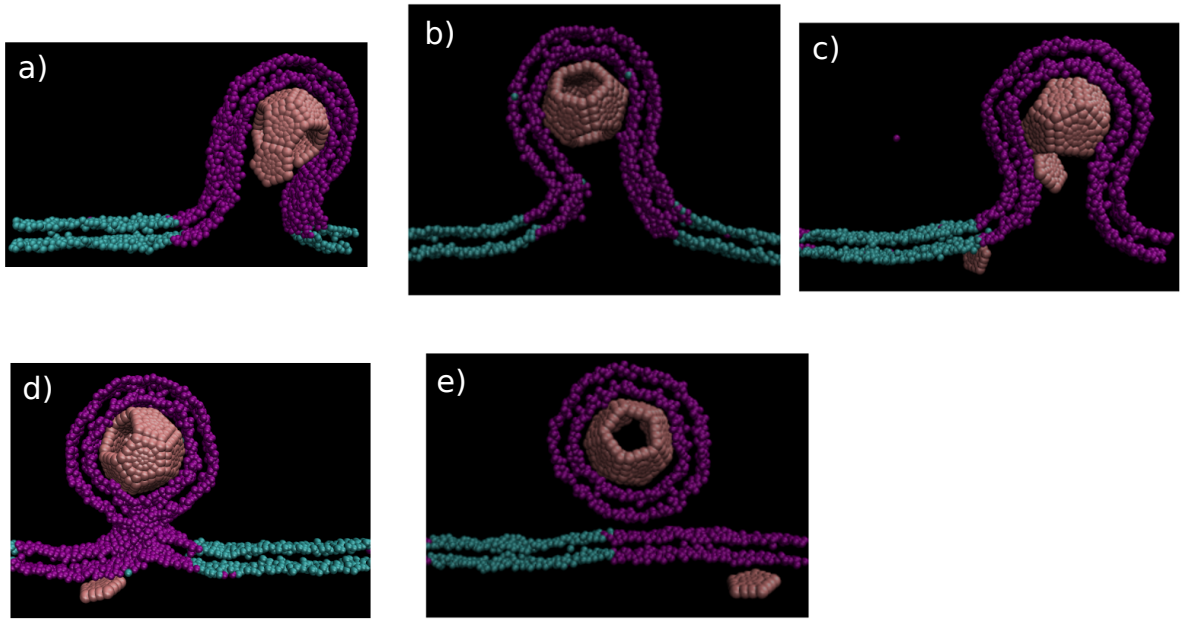

Figure 5.26: Assembly and budding of a capsid with a hole. in a) a pentamer assembles forming a hole. The hole lies inside the area wrapped by the membrane, and when the eleventh particle assembles, the capsid buds.

domain surface. The adhesion of the growing capsid however induces deformations of the order of the capsid size. When the interacting domain size is similar to that of the capsid, the curvature produced in the domain will have a correlation length similar to the domain diameter, and will take the whole domain out of the membrane plane and trigger budding. In bigger domains however, with sizes large compared to the the capsid size, the curvature produced by the growing capsid will only produce small protrusions on the domain surface that will not result in the bending of the whole domain, so it will remain at equilibrium in the membrane plane.

In general, for small line tensions, and even when the whole domain does not bud, the

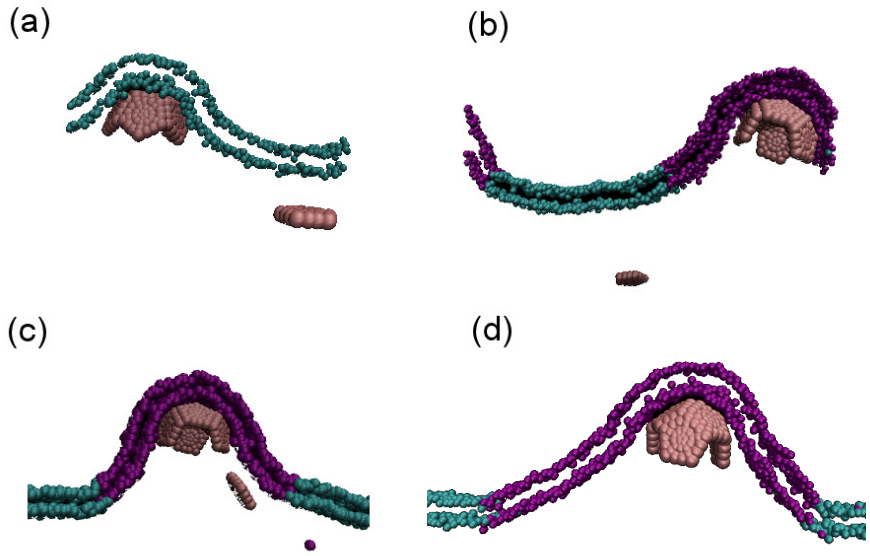

Figure 5.27: Domain effect on the neck and kink shapes in half capsids with $\epsilon_{DM} = 0.85\epsilon_0$ and $\epsilon_s = 0.4\epsilon_0$. The domain is bigger in each figure. a) Membrane with size 60x60σ_0^2 with no domain, seventh particle never assembles. b) small domain in medium membrane, membrane size 100x100σ_0^2, domain radius $R = 30\sigma_0$. The seventh particle attaches, but the domain is too small to cover the full capsid, so budding occurs through membrane rupture. c) medium membrane and medium domain with radius $R = 35\sigma_0$, for this set of parameters the whole domain buds with partial capsid of 8 subunits. d) medium membrane and large domain with radius $R = 40\sigma_0$, for this conditions, we found right budding for half of the simulations and domain budding for the other.

bending induced by the growing capsid equilibrates the length of the border at some new value. This equilibration is usually related with the appearance of long necks that are compatible with the decrease of the border length and with smooth shapes of small curvature. The length of the neck and the curvature degree depend on the size of the domain (Fig. 5.27) and on the adhesion strength between the capsomers and the membrane. As we saw in the study of the wrapping of the spherical particle (Sec. 4.4.2), high adhesions are related to tight wrapping, with high bending areas in the kink region due to the maximization of the membrane contact with the capsid, whereas small adhesion strengths are responsible of configurations where the bending is reduced by the formation of long necks and soft kinks.

The soft kinks on the membrane around the growing capsid make easier the attachment of a new capsomer in two ways. First, the particle diffuses easily through smooth kinks towards the growing capsid, and second, the work done by the particle to bend the membrane when assembling is reduced because the conformational change of the membrane is smaller (Fig.

5.17)

5.5.4 Budding at very low capsomer concentration

As a first approach to study the simultaneous assembly and budding on a domain for the relevant parameters, we analyze a system at a very low capsomer concentration. Free subunits are injected one by one, as if they arrived to the membrane from the cytoplasm at a very low rate. Only when a particle has assembled into the growing capsid, a new one is injected. For an interacting domain with $r_{\text{domain}} = 35\sigma$ and line tension defined by $\epsilon_{\text{DM}} = 0.85\epsilon_0$, we perform the simulations for different values of the adhesion strength between the membrane and the capsomers and for different line tensions and domain sizes.

As it was commented above, the strength of the line tension and the size of the domain will determine the domain dynamics and its equilibrium configuration. Varying these parameters we observe different assembly products that are summarized in the phase diagrams in Figs. 5.28 and 5.29. In the first one, the behavior of the system as a function of the line tension and the adhesion strength is represented. Six different types of behaviors are observed.

The phase diagram in Fig. 5.28 can be compared to the system behavior predicted by the energy estimators in Fig. 5.14. The region of assembly and budding found with the energy estimators corresponds to the central area of the phase diagram in Fig. 5.28. In the simulations however, two different possible outcomes are observed for that region. For small line tensions we find a broad area where right assembly and budding takes place (green stars), for higher line tensions the whole domain buds before the capsid is complete (yellow squares). As the adhesion strength increases the high curvatures adopted by the domain around the capsid prevent the budding of the whole domain, so the number of capsomers found in the final configuration increases. In the case where a complete capsid buds, the domain wraps the growing capsid forming intermediate necks that do not promote the budding of the whole domain but produce soft kinks that facilitate the assembly of the last capsomers. In these cases, a complete capsid with icosahedral symmetry stays connected to the rest of the domain trough a think neck until it fuses and the enveloped capsid is released.

For small line tensions (high values of ϵ_{DM}) and small adhesion strengths, the capsid assembles but the membrane does not wrap it (orange triangles), for slightly higher adhesion strengths, the domain wraps the capsid but the gain in adhesion is not enough for budding to proceed, and the assembled capsid stays attached to the membrane through a neck (light blue square). Those behaviors are predicted with the energy estimators and correspond to the left region on Fig. 5.14.

For high adhesion strengths the membrane wraps the growing capsid but there is either no budding (red circles) or the capsid does not assembly correctly (violet circles), this last case corresponds to the behavior on the right area in Fig. 5.14, when the gain upon adhesion is more important than that on assembly, so the capsomers attach to the membrane but do

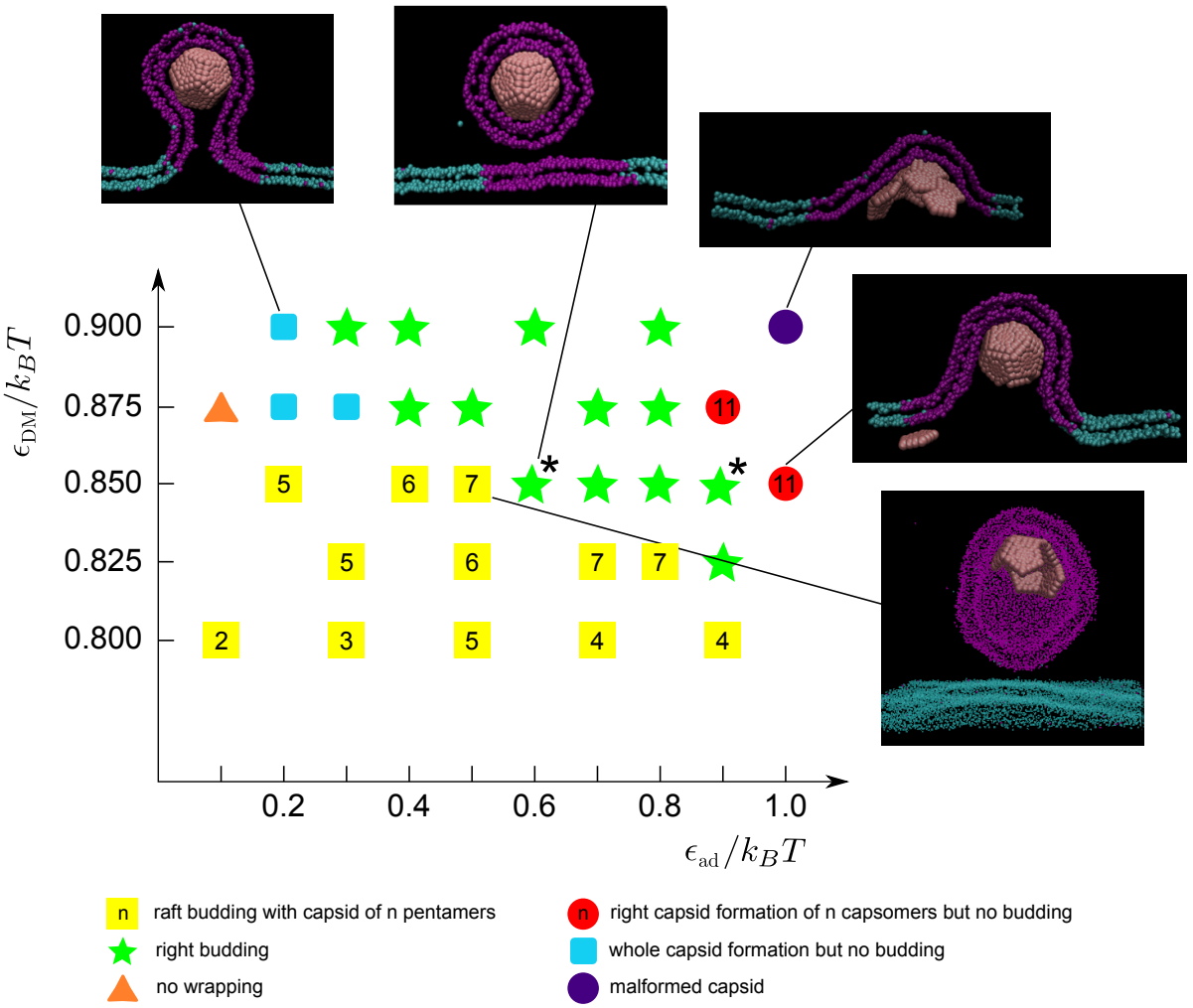

Figure 5.28: Assembly and budding as a function of the interaction between domain and membrane molecules (ϵ_{DM}) and the adhesion strength.

not form right capsids. The cases when the capsomers assemble correctly but budding does not occur are due to a tight wrapping of the membrane that produces short necks with acute kinks that difficult the assembly of new capsomers as explained in Sec. 5.5.3.

The effect of the domain radius on the system behavior is showed in Fig. 5.29. As we expected, the curvature produce by the attachment of the growing capsid promotes budding of the whole domain when it is small. When it is large however, intermediate adhesion strengths that allow budding for smaller domains induce curvatures only on small fractions of the domain so long necks are not promoted. The capsid is tightly wrapped and the surrounding membrane forms acute kinks that prevent the last particles to assemble, so the equilibrium configurations corresponds to an unfinished capsid wrapped by a small fraction of the domain.

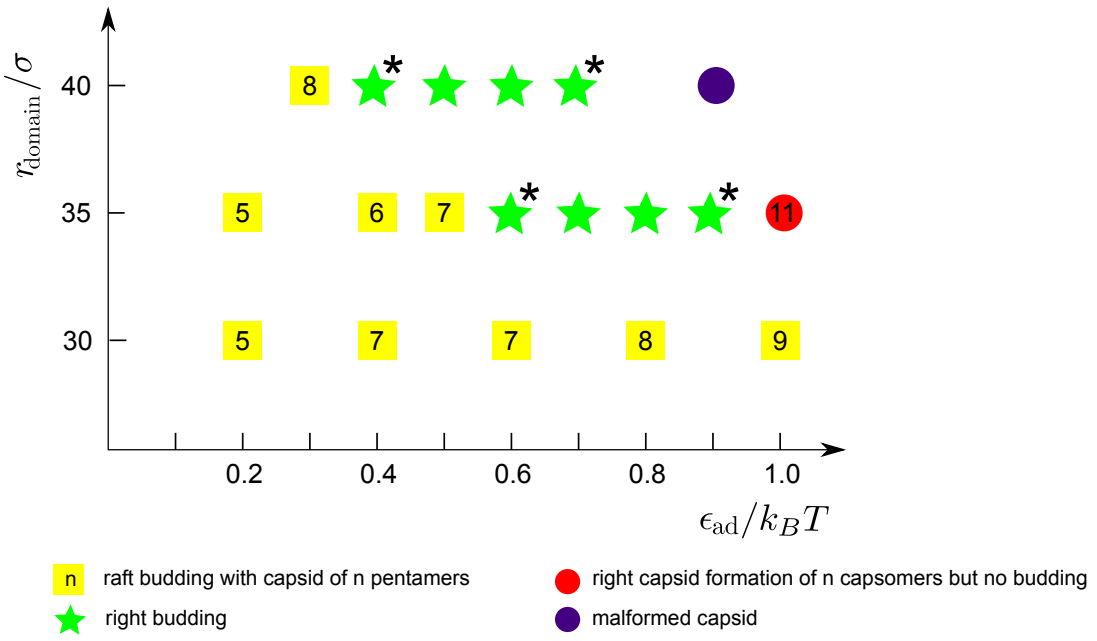

Figure 5.29: Assembly and budding as a function of the domain radius and the adhesion strength for a domain with $\epsilon_{\mathrm{DM}} = 0.85$.

5.5.5 Budding dynamics

When pentamers are injected one by one, the capsid grows through quasi-equilibrium states and the process is fully determined by the dynamics of the domain, the typical time for curvature creation and, if applies, for budding. The rate at which pentamers find each other at the domain can determine the extent at which the capsid can grow. This rate would be related to the pentamer concentration. To investigate this relation, budding is studied for different injection rates and the results are summarized in a phase diagram of the system behavior as a function of the adhesion strength and the injection rate (Fig. 5.30).

For $\epsilon_{\mathrm{ad}} = 0.6\epsilon_0$ complete assembly and budding occurs for any injection rate. For the rest values of the adhesion strength however, the system behavior depends strongly on the injection rate. Low injection rates allow assembly and budding for high adhesion energies because the capsomers have time to assemble correctly before new monomers arrive. High injection rates in turn prevent the domain to bud before the capsid is fully formed, because capsid assembly occurs faster than domain budding.

In general, low injection rates behave similar to the quasi-equilibrium case. Intermediate to high injection rates however give rise to kinetic traps for high adhesion energies but allow right capsid assembly and budding at low adhesion energies.

For intermediate rates with not so high adhesion strengths, the system is trapped by

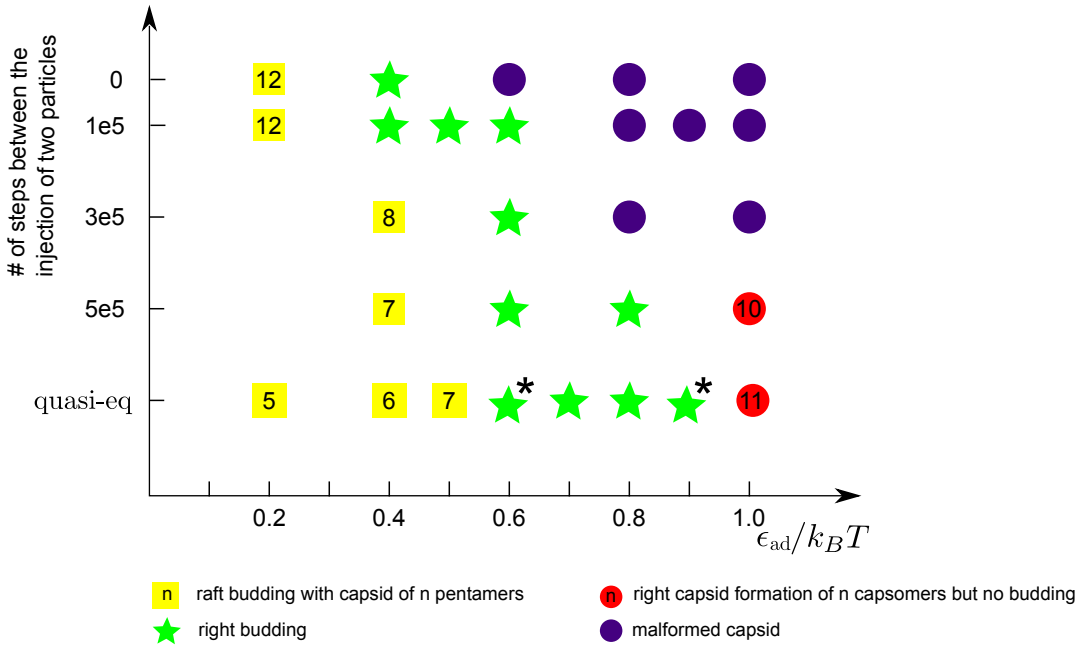

Figure 5.30: Assembly and budding as a function of the injection rate and the adhesion strength.

monomer starvation, two partial capsids are usually formed that do not assemble into a single one (Fig. 5.31), and no free monomers are available for further assembly. The capsomers mobility is very slow at these adhesion strengths so disassembly and new assembly are not observed. When the two partial capsids find each other, they usually fused in a bigger aggregate without icosahedral symmetry. For very high adhesion strengths, malformed capsids appear when a pentamer does not bind correctly to the growing capsid preventing the rest of pentamers to assemble correctly (Fig. 5.32).

The same kinetic traps that arise for high subunit binding energies or for high concentrations are observed here as a function of the adhesion strength.

At very high injection rates most of the pentamers attach to the domain before assembling. Only for small adhesion strengths, correct assembly proceeds from crowded environments (Fig. 5.33), for higher adhesion strengths however, capsomers are not usually able to assemble into well formed capsids. The presence of the many free capsomers inhibit the curvature enhance from the line tension, so the capsids cannot assemble. In these cases, big flat aggregates that do not bend the membrane (Fig. 5.34), or half capsids emerging of a flat aggregate constituted by the rest of the pentamers (Fig. 5.35) are usually observed.

When more than 12 capsomers are injected, the subunit concentration on the domain surface increases considerably and for some conditions the typical kinetic traps arise. However, for an intermediate value of the adhesion strength, where capsomer mobility is not highly

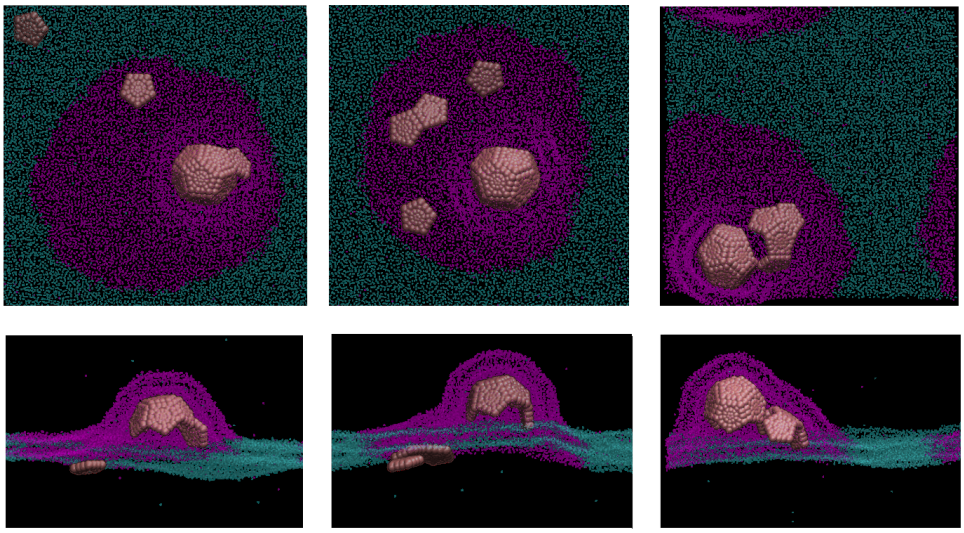

Figure 5.31: Top and side views of a system with intermediate injection rate and adhesion strength. Two partial capsids form.

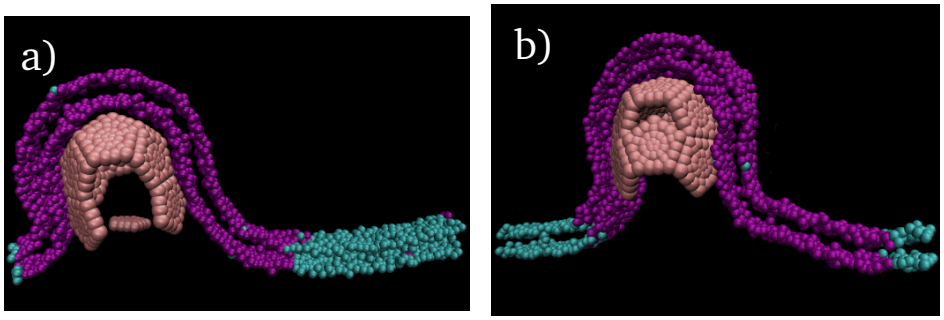

Figure 5.32: Consequences of wrong pentamer assembly. A)Wrong assembly of two pentamers that bind to the capsid but do not assembly correctly. b) The next cannot assembly in the right position.

reduced, partial capsids are formed. The subunits that do not belong to the capsid usually constitute a flat aggregate that difficult budding of the growing capsid, however, for low adhesion strengths with high pentamer mobility, budding has been observed in the presence of 19 capsomers, near 60% more than needed for capsid formation (Fig. 5.36)

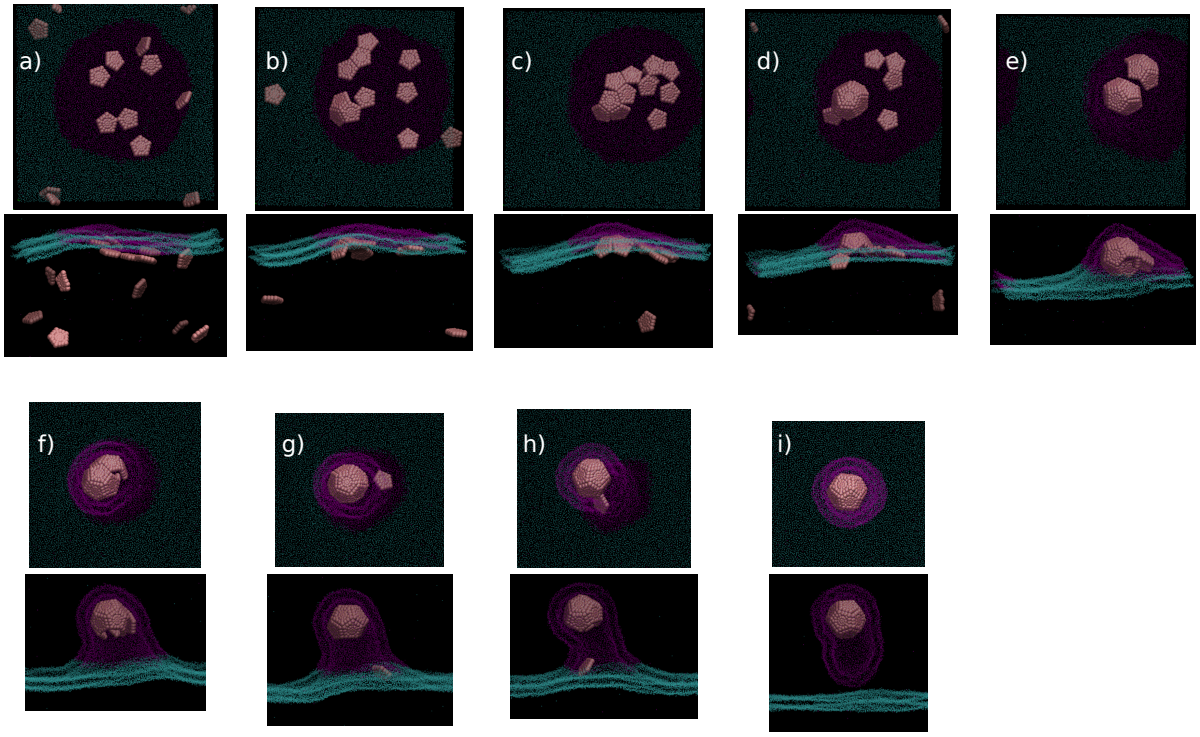

Figure 5.33: Top and side views of assembly of 12 free particles injected all together for a small adhesion strength. In this regime the pentamers show high mobility and ability to accommodate in the growing capsid. In e) the assembly of two big aggregates (with 8 and 4 capsomers) shows this high adaptation. Finally the whole domain buds when the capsid is formed

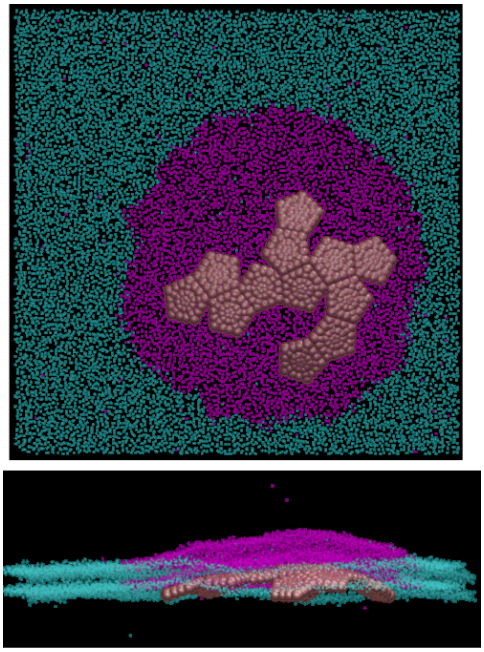

Figure 5.34: Consequences of a crowded environment. For very high adhesions and high injection rates, the capsomers attach to the membrane but do not assemble and stay in a flat aggregate

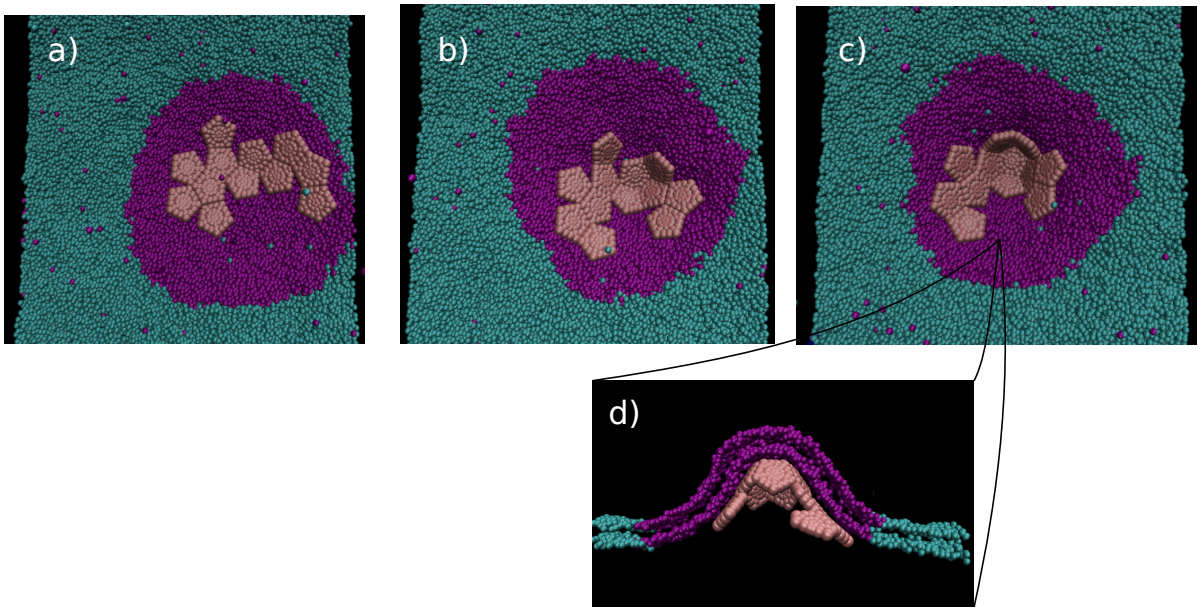

Figure 5.35: Consequences of a crowded environment. For high adhesions (not as high as in Fig. 5.34) and high injection rates, some capsomers assemble into a half capsid, the rest of capsomers form a flag aggregate around it that eliminate the curvature enhancement produced by the domain.

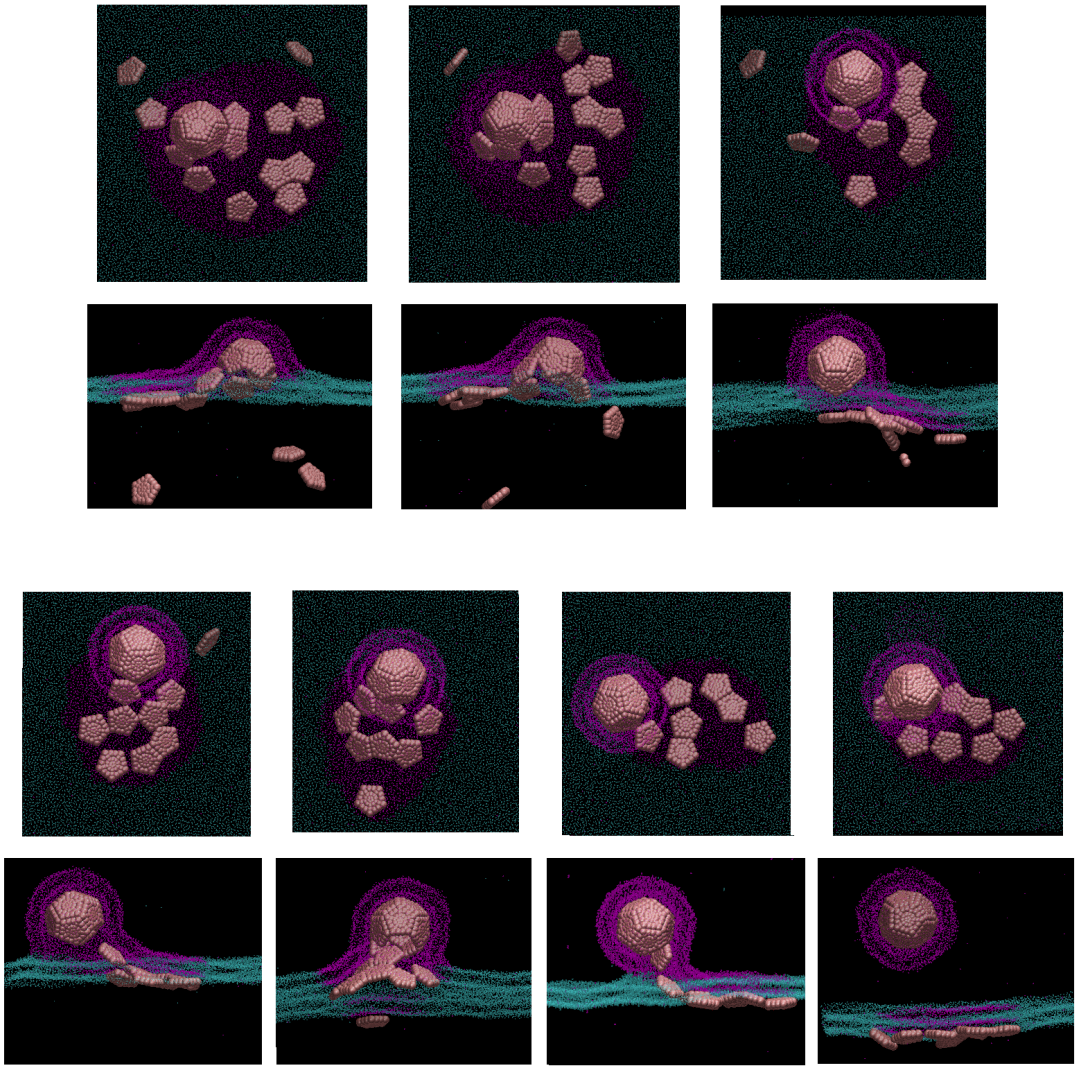

Figure 5.36: Assembly and budding with 20 capsomers. Top and side views of the process.

5.6 Discussion

Our study of a simplified model for simultaneous assembly and budding of a virus capsid on a domain allowed us to analyze the requirements for capsomer assembling on a membrane, and the behavior of the system as a function of the domain properties and the adhesion strength.

We have seen that assembling on the membrane has specific constraints that impose certain characteristics on the capsid model that do not apply to the bulk case. For capsids to assemble correctly on the membrane, the model needs to maximize the capsomer polar specificity while reducing the azimuthal specificity. A high polar specificity promotes structures with the right symmetry, while a reduced azimuthal specificity increases orientational flexibility of interacting capsomers, so that the binding of subunits is not highly constraint by the local curvature of the membrane.

The requirements on angle specificity together with a carefully modelization of the adhesion between the capsomers and the membrane led us to modify Wales Model [180] to account for the special features of assembly in a twodimensional surface. The adhesion between the capsomer and the membrane has been modeled via a set of interacting points on the capsomer body that interact with the hydrophilic core of the membrane, mimicking the action of matrix proteins. The membrane in turn has been represented using a coarse grained model that retains the main properties of a lipid bilayer, while reducing importantly the degrees of freedom of the system, so that dynamics can be efficiently investigated for large length and time scales.

Virus assembly during budding involves the binding of capsomers to the membrane at the same time that they assemble to form a new capsid. These two processes lead to the bending of the surface while it wraps the growing capsid, and are coupled by the capsomer adhesion to the membrane. We have studied how curvature needs to be created on the membrane for the capsomers to assemble. In our model of 12 pentamers, assembly and adhesion can not drive the changes in curvature needed for formation of a complete capsid and budding, so the process rely on extra elements that induce membrane curvature. This is compatible with the actual knowledge on virus budding that recognizes the action of cytoskeleton [39, 214], matrix proteins [69] or other cellular mechanisms in the creation of membrane curvature [34, 39, 192].

Following the general idea supporting the role of membrane microdomains as budding sites [30, 42], in our study we have identified the budding site as a microdomain with tunable properties as size or interfacial energy. We have analyzed the spontaneous assembly and budding of capsomers on microdomains, and we have showed that domains can induce the required curvature on the membrane for capsomer assembly in our model. The dynamics of the domain alone have also been investigated, and the values for line tension and size at which it can bud only driven by thermal fluctuations have been obtained.

From the energy estimation of the hypothetical equilibrium states after budding we have identified three possible behaviors depending on the balance between the bending, adhesion and assembly energy: assembly and budding, no wrapping and adhesion without capsid formation. The quasi equilibrium simulations at very low injection rate however show that the system dynamics give rise to certain subdivisions of the general predicted behaviors, resulting in 6 different types of outcomes. Wrapping and budding cases are divided into those where the whole domain buds before the capsid is complete and the cases where only the enveloped particle buds after assembly. Adhesion without assembly happens in two different ways, either malformed particles or incomplete capsids showing icosahedral symmetry are obtained, but none of them result in budding. Finally, the no wrapping behavior is constituted by the cases where a partial capsid barely attaches to the membrane, and the cases where the membrane wraps the capsid but budding does not occur.

The results obtained by Briggs, Carlson and coworkers on HIV budding [64, 65] where wild virions bud before the capsid is complete suggesting that the ESCRT-mediated release would occur concomitant and in kinetic competition with capsid assembly, are similar to the cases we find where the whole domain buds before capsid completion. This confirms the fact that the domain interfacial energy favors curvature enhancemente and budding in a similar way that other cell machinery components. On the other hand, the modified virions lacking the ability to exploit he ESCRT complex that assembled into a whole capsid but did not detach from the membrane, would be similar to the cases where the whole capsid stays attached by a thin tether for small interfacial and low adhesion energy. Parameters that allow right assembly and budding could represent other type of viruses that bud with complete capsids without cellular machinery or at least, where this machinery only acts at the final stage of the budding process and is not in kinetic competition with assembly. Finally, malformed capsids or uncompleted capsids that stays attached to the membrane may possibly represent system conditions beyond physiological values of the adhesion energy at which the virus cannot complete its life cycle.

As observed by Matthews and coworkers [224], we find that assembly on the membrane is possible for a concentration that does not allow assembly in bulk. The membrane acts as a concentration device, so assembly only depends on the surface concentration. This mechanism allow low capsomer concentration in the cytosol, so that the capsomers can only assemble while attached to the surface. Although it has been argued that the matrix and transmembrane glycoproteins may have a role triggering and stabilizing capsid assembly [60], these results confirm a more basic role of the membrane as a concentration device reducing the amount of capsomers that need to be created by the cell to achieve the optimal concentration for assembly.

The effect of the injection rate has also been studied to analyze how assembly and budding depend on the rate at which capsomers arrive to the budding site. We have seen that in general, for low injection rates the system behaves similar to the quasi-equilibrium case.

Intermediate to high injection rates however give rise to kinetic traps for high adhesion energies but allow right capsid assembly and budding at low adhesion energies. The same kinetic traps that arise for high subunit binding energies or for high concentrations are observed here as a function of the adhesion strength. From Ivanchenko results [220], we can argue that the best modelization of the assembly process would be for an intermediate to low injection rate at which the capsid is formed simultaneously with capsomer arrival to the budding site, so big flat aggregates forming when the capsomers arrive faster than assembly takes place would not represent the natural process.

Finally, we should discuss the validity of our model based on the capsid structure and size. The diameter of our T=1 assembled capsid is around 17 nm, while the effective diameter of the enveloped particle is around 22 nm. The smallest enveloped viruses that follow this route for assembly and budding are those from flariviridae family with a nucleocapsid diameter around 30 nm and a total diameter around 40-50 nm and present a T=3 or T=4 structure [227]. From the previous work on Chapter 4 we now that the effect on the adhesion and bending energy on budding of capsids bigger than 18 nm is not very strong (Fig.4.6). However, the small size of the particle in our model compared to typical enveloped virus can be determinant on the limitations of the model on assembly and curvature induction on the membrane. A bigger capsid constituted by a higher number of capsomers would be able to induce strongest changes in the membrane curvature. Moreover, since the bending energy of a vesicle does not depend on the size, the work done by a single capsomer to bend the surface in a model with a higher number of subunits would be considerably smaller. For this reason, it would be interesting to study the effect of a T=3 capsid on the already analyzed system behavior.

General discussion

The results presented in this thesis provide new insights into the dynamics of virus budding. The fundamental physical interactions that control passive budding in systems where assembly and budding occur sequentially, and systems in which budding and assembly are concomitant, are identified and the resulting dynamics are thoroughly analyzed. The results obtained in this work are of special interest because they provide a quantitative description that elucidates the minimum requirements for virus assembly and budding. In an experiment-driven field where it is difficult to distinguish the contributions of individual factors that allow budding, physical modeling of the system gives important insights in the fundamental mechanisms of the process.

One of the fundamental issues we address in this thesis is the determination of the extent at which minimum models can provide a good representation of a biological complex system. Minimum models and coarse grained approximations are not novel approaches to these kind of systems, and however they are somehow undermined by atomistic models that produce high detailed descriptions and can be used to accurately quantify specific magnitudes. The profits of the atomistic approach are evident, and their limitations on representing macroscopic systems that require extremely long computation times may appear prohibited today, but it could be only a question of time that computers are fast enough so that very complex systems can be represented with atomistic detail. The results obtained in this thesis support the use of minimum models for other reasons than simply overcoming computational limitations. As we have seen in this work, these models are able to reproduce the essential properties of a system, and therefore, they provide a powerful conceptual tool for understanding the basic features and for distinguishing the main factors that are responsible of a process. Although very often these kind of models that allow to span the microscopic and macroscopic scales are presented mainly as methodological tools to deal with the limitations of atomistic models, they in fact represent a different approach that is not able to focus accurately on specific features but allow to understand the complexity of the system from a simplified global view.

Within this context of minimum models, we understand the results obtained here as a description of the basic factors controlling virus assembly and budding. Using elastic theory and molecular dynamics simulations, we have analyzed in the first part of the work the process where a nucleocapsid assembles before budding, and in the second part, the case where assembly and budding occur simultaneously. In both cases we have verified that the

existence of attractive interactions between a particle and the phospholipids, which has been proposed as the minimal requirement for viral budding is sufficient to drive efficient wrapping. We have also determined the behavior of the system and discussed the outcomes from a biological perspective.

To study systems where assembly precedes budding, we have considered a minimal model in which the membrane is treated as a bilayer of homogeneous composition, and the pre-assembled capsid is represented by a spherical particle. Using molecular dynamics simulations four classes of behaviors resulting from the interaction of a particle with a membrane have been identified: no wrapping or adhesion resulting in a minimal perturbation of membrane configurations, partial wrapping, complete wrapping, and wrapping via rupture of the membrane. Equilibrium calculations showed that there are only two equilibrium configurations, corresponding to no wrapping or complete wrapping. The primary difference between the elastic model and the finite-time dynamical simulation results is the appearance of long-lived partial wrapping states. These states had been observed not analyzed before. We studied the free energy along the penetration, which we found was a good order parameter to describe the process, and identified those partial wrapping states as metastable. These results mean that the system is confined to a relatively narrow range of adhesion strengths to avoid stalled partially-wrapping states and membrane rupture. While this result is qualitative, it does establish important constraints on viral evolution if budding is limited to these ingredients.

Simultaneous assembly and budding was analyzed similarly to the previous case but using a simplified model of virus assembly instead of the spherical particle. The adhesion between the capsomers and the bilayer was represented by an interaction that reproduces the role of a matrix protein. Studies of assembly and budding with a model of this realism had not been applied to this process before. We used molecular dynamics and elastic theory on this model to analyze two different problems. First we determined the requirements for capsid models for assembly on a membrane different to those that apply for assembly in bulk, and secondly we studied the behavior of the system when the capsomers assemble on a micro-domain, distinguishing between the analysis of quasi-equilibrium states where just one capsomer at a time assembles into the growing capsid, and the study of the dynamical case where capsomers arrive to the budding site at a constant frequency.

We have observed that assembly on the membrane has specific constraints that impose certain characteristics on the capsid model that do not apply to the bulk case. The membrane imposes some requirements on the orientational specificity of the capsomer-capsomer interaction by restricting the angles of associating capsomers. For capsids to assemble correctly on the membrane, the model must increase orientational flexibility of interacting capsomers, so that the binding of subunits is not highly constraint by the local curvature of the membrane, while enhancing assembly on the membrane surface with icosahedral order even for low capsomer-capsomer binding energy. Another important requirement for assembly on a

membrane is the creation of curvature on the membrane around the growing capsid. We have seen that in our model of T=1, with 12 pentamers, assembly and adhesion can not drive the changes in curvature needed for formation of a complete capsid and budding, so the process relies on extra elements that induce membrane curvature. We found that a domain can make this function so right assembly and budding are promoted on its surface. Finally, we also observed that the membrane acts as a concentration device, so assembly only depends on the surface concentration, and it is possible for concentrations much lower than those needed for bulk assembly. This mechanism allows low capsomer concentration in the cytosol, so that capsomers can only assemble while attached to the surface, reducing the amount of capsomers that need to be created by the cell to achieve the optimal concentration.

The quasi-equilibrium analysis allowed us to identify the possible outcomes depending on the adhesion energy and the domain line tension. From the energy estimation of the hypothetical equilibrium states after budding we have identified three possible behaviors: assembly and budding, no wrapping, and adhesion without capsid formation. The quasi equilibrium simulations at very low injection rate however show that the system dynamics give rise to certain subdivisions of the general predicted behaviors, resulting in 6 different types of outcomes. Wrapping and budding cases are divided into those where the whole domain buds before the capsid is complete and the cases where only the enveloped particle buds after assembly. Adhesion without assembly happens in two different ways, either malformed particles or incomplete capsids showing icosahedral symmetry are obtained, but none of them result in budding. Finally, the no wrapping behavior is constituted by the cases where a partial capsid barely attaches to the membrane, and the cases where the membrane wraps the capsid but budding does not occur.

The possible behaviors we found are easily related to previous experimental results. The cases where the whole domain buds before capsid completion show similar properties to the ESCRT-mediated release in HIV budding where wild virions bud before the capsid is finished. While the cases where the whole capsid stays attached by a thin tether would resemble those of mutant HIV virions lacking the ability to exploit he ESCRT complex, that assembled into a whole capsid but did not detach from the membrane. These similarities confirm the role of a micro-domain as a budding site that induces assembly and budding in a similar way than cellular machinery that assists budding. Finally, malformed capsids or uncompleted capsids that stay attached to the membrane may possibly represent system conditions beyond physiological values of the adhesion energy at which the virus cannot complete its life cycle.

The effect of the injection rate has also been studied to analyze how assembly and budding depend on the rate at which capsomers arrive to the budding site. We have seen that, as proposed, the highest efficiency occurs at a low rates, where the capsid is formed simultaneously with capsomer arrival to the budding site. On the other hand, intermediate to high injection rates give rise to the same kinetic traps that arise for high subunit binding energies

or for high concentrations.

The minimum models used in this work allow to identified the basic properties of assembly and budding, but we should understand their limitations on the description of real processes. Depending on the viral system, a number of additional phenomena contribute to budding, including membrane-associated viral envelope or spike proteins, preferential budding from lipid rafts, the use of cytoskeletal machinery to actively drive or assist budding or scission, and the ability of the virus to remodel cell membrane properties. Our models do not account for these effects that can broaden the range of functional adhesion energies that enable efficient viral egress.

Also, it is important to evaluate the differences in size and geometry of virus capsids compared to those used in our study. In the first part of the work, the particle diameters in our simulations range from 9 to 36 nm. Although this can be small for viral capsids that range from 40 nm (e.g. hepadnavirus) to hundreds of nanometers, we have seen that the results can be extrapolated into that range, because the adhesion-wrapping transition decreases with radius as $1/R$ for constant bending rigidity.

In the second part of the work, the diameter of our T=1 assembled capsid is around 17 nm, while the effective diameter of the enveloped particle is around 22 nm. The smallest enveloped viruses that follow this route for assembly and budding are those from flariviridae family with a nucleocapsid diameter around 30 nm and a total diameter between 40 and 50 nm, and present a T=3 or T=4 structure. As we saw before, the extrapolation could also be applied here, although the small size of the particle in our model compared to typical enveloped viruses can be determinant on the limitations of the model as for assembly and curvature induction on the membrane. A bigger capsid constituted by a higher number of capsomers would be able to induce strongest changes in the membrane curvature. For this reason, it would be interesting to study the effect of a bigger T=3 capsid.

Conclusiones finales

Los resultados obtenidos en esta tesis proporcionan nuevos conocimientos de la dinámica de la gemación de virus. Se ha descrito ampliamente el comportamiento del proceso y se han identificado las interacciones físicas fundamentales que controlan la gemación pasiva en sistemas donde el autoensamblaje de la cápsida y la gemación ocurren sucesivamente y aquellos en los que ambos prcesos tienen lugar de manera simultánea. Los resultados obtenidos en este trabajo son de especial interés porque proporcionan una descripción cuantitativa que permite determinar los requisitos mínimos del proceso de autoensamblaje y gemación. En un campo principalmente experimental, y en un proceso en el que es dificil distinguir las contribuciones de los factores principales, la modelización del sistema ofrece información importante sobre los mecanismos esenciales de la gemación.

Uno de los temas fundamentales que se tratan en esta tesis es la determinación de la validez de los models de mínimos en la representación de sistemas biológicos complejos. Los modelos de mínimos y las aproximaciones 'coarse grained' no constituyen novedosas formas de describir este tipo de sistemas, pero últimamente han sido desplazadas por estudios atomísticos que permiten una descripción con más detalle y el cálculo con gran precisión de magnitudes específicas. Los beneficios de las representaciones atomísticas son evidentes, y sus limitaciones en la descripción de sistemas complejos que requieren largos tiempos de computación pueden parecer prohibitivos hoy en día, pero es posible que sólo sea cuestión de tiempo que los ordenadores sean suficientemente rápidos y permitan una representación atomística de estos sistemas. Los resultados obtenidos en esta tesis apoyan la validez de los modelos de mínimos por razones diferentes a las limitaciones computacionales. Como hemos visto en este trabajo, estos modelos son capaces de reproducir las propiedades esenciales de un sistema , y por lo tanto constituyen una herramienta conceptual muy potente para entender las caracetrśiticas básicas y distinguir los factores principales responsables de un proceso. Aunque estos modelos se presentan habitualmente como meros recursos metodológicos para conectar las escalas microscópicas y macroscópicas, representan sin embargo un enfoque diferente que no permite la descripción de detalles específicos pero ofrece una visión global y simplificada del sistema que ofrece una comprensión general de sus propiedades básicas.

En este contexto de modelos de mínimos, analizamos los resultados presentados en los anteriores capítulos como una descripción de las características básicas que controlan el autoensamblaje y la gemación de partículas víricas. Mediante teoría elástica y simulaciónes de

dinámica molecular hemos analizado en la primera parte del trabajo el proceso en el que el autoensamblaje de la cápsida precede a la gemación, y en la segunda parte el caso en el que ambos procesos ocurren simultáneamente. En ambos casos hemos confirmado que la existencia de interacciones atractivas entre una partícula vírica y los lípidos de la membrana, que se había identificado anteriormente como requisito mínimo para la gemación, es suficiente para que se dé el proceso. Hemos caracterizado también el comportamiento del sistema y tratado su posible significado biológico.

En el estudio de sistemas donde el autoensamblaje precede a la gemación hemos usado un modelo de mínimos en el que la membrana es tratada como una bicapa con composición homogénea, y la cápsida previamente autoensamblada la representamos con una partícula esférica. Las simulaciones de dinámica molecular nos han permitido identificar cuatro comportamientos diferentes en función de la intensidad de la interacción entre la partículas y la membrana: la partícula no se envuelve o ni siquiera se adhiere a la membrana dando lugar a perturbaciones mínimas de la superficie de la bicapa lipídica, la envoltura parcial de la partícula, la envoltura total y gemación, y gemación por ruptura de la membrana. Los cálculos en equilibrio con teoría elástica han mostrado sin embargo que sólo hay dos tipos de configuraciones de equilibrio que corresponden a los casos en los que la partícula no se envuelve y los que se envuelve completamente y acaba dando lugar a gemación. La diferencia principal entre los resultados obtenidos con el modelo elástico y los de las simulaciones dinámicas a tiempos finitos, es la aparición de estados con vida media larga en los que la partícula aparece envuelta parcialmente. Estos estados habían sido observados anteriormente pero no analizados con detalle. Hemos obtenido la energía libre del sistema a lo largo de la penetración, que ha resultado ser un buen parámetro de orden del proceso, y hemos identificado estos estados como metaestables. Estos resultados indican que el sistema está confinado en un rango de adhesiones relativamente estrecho para evitar estados de envoltura parcial y de ruptura de membrana. Aunque este resultado es cualitativo, establece restricciones importantes en la evolución de los virus si la gemación está limitada a estos factores.

El caso en el que autoensamblaje y gemación ocurren simultáneamente se ha analizado de forma similar al anterior pero usando un modelo simplificado para el autoensamblaje de la cápsida en lugar de la partícula esférica. La adhesión entre la membrana y los capsómeros se ha representado con una interacción que reproduce la función de las proteínas matriz. Estudios de autoensamblaje y gemación de un modelo con este realismo no se habían llevado acabo anteriormente. Hemos usado de nuevo teoría elástica y simulaciones de dinámica molecular para investigar dos aspectos del proceso. Primero hemos determinado los requisitos que tiene que cumplir el modelo de la cápsida para que se produzca autoensamblaje en una superficie en lugar de en volumen, y después, hemos estudiado el comportamiento del sistema cuando los capsomeros se autoensamblan en un microdominio, distinguiendo entre el análisis de los estados de quasi-equilibrio donde sólo un capsómero se ensambla en

la cápsida cada vez, y esl estudio dinámico donde los capsómeros llegan al dominio a una tasa constante.

En el primer caso, hemos observado que el autoensamblaje en la superficie de la membrana tiene ciertas restricciones que imponen que el modelo de la cápsida cumpla unas condiciones que no son necesarias cuando el autoensamblaje tiene lugar en volumen. La membrana impone ciertos requisitos en la especificidad orientacional de la interacción entre capsomeros al restringir los ángulos en los que los capsómeros pueden enlazarse. Un modelo de autoensamblaje que dé lugar a la formación de cápsidas en la superficie de la membrana tiene que tener una alta flexibilidad orientacional en la interacción atractiva entre capsómeros para que esta no esté totalmente limitada por la curvatura local de la membrana, a la vez que fomente la formación de cápsidas con la geometría adecuada incluso para valores bajos de la interacción entre capsómeros. Otro requisito importante para el correcto autoensamblaje en la membrana es la inducción de curvatura alrededor de la cápsida. Hemos visto que en nuestro modelo de T=1 con 12 pentámeros, el autoensamblaje y la adhesión no pueden dar lugar a los cambios de curvatura en la membrana necesarios para que se produzca la gemación, por lo que el proceso depende de elementos extras que induzcan curvatura. Se ha observado que un dominio puede suplir esta función y que es capaz de provocar la gemación de cápsidas con la geometría adecuada que se han autoensamblado en su superficie. Por último, hemos observado que la membrana actúa como una plataforma de concentración de capsómeros, de esta forma el autoensamblaje sólo depende de la concentración superficial y tiene lugar para concentraciones menores que cuando occurre en volumen. Este mecanismo permite una concentración menor de capsómeros en el citosol de manera que sólo pueden autoensamblarse en la membrana, reduciéndose así el número de copias de proteínas de cápsida que tiene que hacer la célula para conseguir la concentración óptima para que se formen las partículas víricas.

Los análisis de quasi-equilibrio nos han permitido identificar los diferentes tipos de comportamientos del sistema en función de la energía de adhesión y del tamaño del dominio y la tensión lineal que experimenta su contorno. De la estimación de la energía de los hipotéticos estados de equilibrio tras la gemación hemos encontrado tres tipos de comportamientos: autoensamblaje y gemación, cápsidas que no son envueltas, y adhesión de capsómeros a la membrana que no forman cápsidas completas. Las simulaciones de los estados de quasi-equilibrio sin embargo muestran que estos tres comportamientos presentan subdivisiones, y señalan 6 tipos de resultados. El autoensamblaje y gemación se divide con los nuevos resultados en los casos en los que el dominio entero froma una vesícula y se separa de la membrana con una cápsida a medio formar en su interior y los que dan lugar a la gemación de una cápsida entera y bien formada envuelta por una porción del dominio. Los casos de adhesión de monómeros que no se autoensamblan en una cápsida tienen lugar de dos formas diferentes en las simulaciones, o bien las cápsidas no se forman correctamente o aun teniendo la geometría adecuada quedan inacabadas, y en ningún caso el proceso acaba con la gemación

de la partícula. Finalmente, el estado en el que la partícula no era envuelta por la membrana muestra también dos variantes: cápsidas parciales que apenas están en contacto con la membrana y cápsidas completas y envueltas pero que no son liberadas y se mantienen conectadas a la bicapa.

Los diferentes comportamientos encontrados en las simulaciones se pueden relacionar con resultados experimentales previos. Los casos en los que el dominio entero forma una vesícula que se separa de la membrana antes de que la cápsida haya terminado de formarse son similares a la gemación mediada por ESCRT de partículas víricas incompletas de VIH. Mientras que los casos en los que la cápsida completa permanece conectada a la membrana son similares a virus de VIH mutados que no son capaces de utilizar el complejo ESCRT. Se ha observado que estas partículas se autoensamblan en una cápsida entera, pero no se desprenden de la membrana. Estas similitudes confirman el papel del micro-dominio como un lugar adecuado para el autoensamblaje y la gemación con capacidad de potenciar el proceso de manera similar a la maquinaria celular que utilizan los virus. Finalmente, las cápsidas malformadas o las incompletas que permanecen unidas a la membrana, puede representar condiciones del sistema más allá de los valores fisiológicos para la energía de adhesión para la cual el virus no puede completar su ciclo vital.

El efecto de la tasa de inyección de capsómeros también se ha estudiado para analizar cómo la velocidad a la que llegan los capsómeros al dominio afecta al comportamiento del sistema. Hemos visto que, como se ha propuesto anteriormente, la mayor eficacia se da para tasas bajas, cuando se forma la cápsida simultáneamente con la llegada del capsómero. Por otra parte, las tasas de inyección intermedias a altas dan lugar a las mismas trampas cinéticas que surgen en el autoesamblaje en volumen para energías de enlace entre capsómeros altas o para altas concentraciones.

Los modelos de mínimos utilizados en este trabajo permiten identificar las propiedades básicas del autoensamblaje y gemación de viruss, pero debemos entender sus limitaciones en la descripción de procesos reales. Dependiendo del sistema viríco, un número de fenómenos adicionales contribuyen a la gemación, incluyendo la acción de diferentes tipos de proteínas de membrana, la composición y características del dominio, el uso de maquinaria celular o del citoesqueleto para producir o ayudar activamente en la envoltura de la partícula, en la gemación o en la escisión final, y las propiedades de remodelación de la membrana celular. Nuestros modelos no tienen en cuenta estos efectos que pueden ampliar el rango de energías de adhesión que permiten la salida del virus de manera eficiente.

También es importante evaluar las diferencias entre el tamaño y la geometría de cápsidas víricas y de las partículas utilizadas en nuestro estudio. En la primera parte del trabajo, los diámetros de la partícula en nuestras simulaciones se encuentran entre 9 y 36 nm. Aunque este puede parecer pequeño en comparación con las cápsidas virícas que se encuentran en la naturaleza que van desde los 40 nm (por ejemplo, hepadnavirus) a cientos de nanómetros, hemos visto que los resultados se pueden extrapolar a este intervalo, debido a que la transi-

ción entre el estado de envoltura y de sólo adhesión disminuye con el radio de forma $1/R$ para una rigidez de membrana constante.

En la segunda parte del estudio, el diámetro de nuestra cápsida es del orden de 17 nm, mientras que el diámetro efectivo de la partícula envuelta es del orden de 22 nm. Los virus con envoltura lipídica más pequeños que siguen esta ruta para la gemación son aquellos de la familia flariviridae con un diámetro de alrededor de 30 nm para la nucleocápsida y de entre 40 y 50 nm para la partícula con la envoltura, y presentar una estructura de T=3 o 4. Como hemos visto previamente, la extrapolación a radios mayores se podría aplicar también en este caso, aunque el pequeño tamaño de la partícula en nuestro modelo en comparación con los virus encapsulados típicos puede ser determinante en las limitaciones del modelo en cuanto al autoensamblaje en membrana y la creación de curvatura. Una cápsida mayor constituida por un número más alto de capsómeros sería capaz de inducir cambios más fuertes en la curvatura de la membrana. Por esta razón, sería interesante estudiar el efecto de una cápsida de T=3 mayor.

Bibliography

[1] S. J. Singer and G. L. Nicolson. The Fluid Mosaic Model of the Structure of Cell Membranes. Science **175**, 720 (1972). (cited on p. 1)

[2] G. M. Cooper and R. E. Hausman. *The Cell: A Molecular Approach* (Sinauer Associates, Inc, 2008), 4th edition. (cited on pages 2, 4, 6, and 7)

[3] J. F. Nagle and S. Tristram-Nagle. Structure of lipid bilayers. Biochim Biophys Acta **1469**, 159 (2000). (cited on pages 2 and 3)

[4] B. A. Lewis and D. M. Engelman. Lipid Bilayer Thickness Varies Linearly with Acyl Chain Length in Fluid Phosphatidylcholine Vesicles. J. Mol. Biol. **166**, 211 (1983). (cited on p. 2)

[5] W. Rawicz, K. C. Olbrich, T. Mcintosh, D. Needham, and E. Evans. Effect of Chain Length and Unsaturation on Elasticity of Lipid Bilayers. Biophysical Journal **79**, 328 (2000). (cited on p. 2)

[6] D. Marsh. Polarity and permeation profiles in lipid membranes. PNAS **98**, 7777 (2001). (cited on p. 2)

[7] J. M. Berg, J. L. Tymoczko, and L. Stryer. *Biochemistry* (W.H. Freeman and Company, New York, 2007), 6th edition. (cited on pages 3, 5, 12, and 13)

[8] R. Lipowsky and E. Sackmann. *Structure and Dynamics of Membranes* (Elsevier, Amsterdam, 2004), 2nd edition. (cited on pages 5 and 13)

[9] G. van Meer, D. R. Voelker, and G. W. Feigenson. Membrane lipids: where they are and how they behave. Nature Reviews Molecular Cell Biology **9**, 112 (2009). (cited on p. 5)

[10] D. R. Critchley, S. Ansell, and S. Dilks. Glycolipids: a Class of Membrane Receptors. Biochemical Society Transactions 1189–1192 (1979). (cited on p. 6)

[11] G. M. Whitesides and M. Boncheva. Beyond molecules: Self-assembly of mesoscopic and macroscopic components. Proc. Nat. Acad. Sci. **99**, 4769 (2002). (cited on p. 8)

[12] C. Tanford. The Hydrophobic Effect and the Organization of Living Matter. Science **200**, 1012 (1978). (cited on p. 8)

[13] S. A. Safran. *Statistical Thermodynamics of Surfaces, Interfaces, and Membranes* (Westview Press, Cambridge, MA, 2003). (cited on pages 9 and 60)

[14] F. H. Stillinger. Water Revisited. Advancement Of Science **209**, 451 (1980). (cited on p. 9)

[15] W. L. Jorgensen and J. D. Madura. Temperature and Size Dependence for Monte Carlo Simulations of TIP4P Water. Mol. Phys. **56** (1985). (cited on p. 9)

[16] D. Chandler. Interfaces and the driving force of hydrophobic assembly. Nature **437**, 640 (2005). (cited on p. 10)

[17] N. T. Southall, K. A. Dill, and A. D. J. Haymet. A View of the Hydrophobic Effect. J. Phys. Chem. B **106**, 521 (2002). (cited on p. 10)

[18] J. N. Israelachvili. Physics of Amphiphiles: Micelles, Vesicles and Microemulsions. In *Physics of Amphiphiles: Micelles, Vesicles and Microemulsions*, 24–58 (North-Holland Physics Publishing, 1985). (cited on p. 11)

[19] W. L. Vaz, F. Goodsaid-Zalduondo, and K. Jacobson. Lateral diffusion of lipids and proteins in bilayer membranes. FEBS Letters **174**, 199 (1984). (cited on p. 12)

[20] D. E. Koppel, M. P. Sheetz, and M. Schindler. Matrix control of protein diffusion in biological membranes. PNAS **78**, 3576 (1981). (cited on p. 12)

[21] R. D. Kornbergt and H. M. Mcconnell. Inside-Outside Transitions of Phospholipids in Vesicle Membranes*. Biochemistry **10**, 1111 (1971). (cited on p. 12)

[22] J. Janiak, M. Small, and G. Graham. Temperature and Compositional Dependence of the Structure of Hydrated Dimyristoyl Lecithin. Biological Chemistry **254**, 6068 (1979). (cited on p. 12)

[23] S. Tristram-Nagle and J. F. Nagle. Lipid bilayers: thermodynamics, structure, fluctuations, and interactions. Chemistry and Physics of Lipids **127**, 3 (2004). (cited on p. 12)

[24] L. Chen, M. L. Johnson, and R. L. Biltonen. A Macroscopic Description of Lipid Bilayer Phase Transitions of Mixed-Chain Phosphatidylcholines: Chain-Length and Chain-Asymmetry Dependence. Biophysical Journal **80**, 254 (2001). (cited on p. 12)

[25] J. Štrancar and Z. Arsov. *Advances in Planar Lipid Bilayers and Liposomes Volume 6*, volume 6 of *Advances in Planar Lipid Bilayers and Liposomes* (Elsevier, 2008). (cited on p. 13)

[26] M. Eeman and M. Deleu. From biological membranes to biomimetic model membranes. Biotechnologie, Agronomie, Société et Environnement **14** (2010). (cited on p. 13)

[27] M. Kranenburg and B. Smit. Phase behavior of model lipid bilayers. The journal of physical chemistry. B **109**, 6553 (2005). (cited on pages 14 and 39)

[28] D. A. Brown and E. London. Structure and Origin of Ordered Lipid Domains in Biological Membranes. Journal of Membrane Biology **164**, 103 (1998). (cited on p. 14)

[29] D. Brown. Structure and function of membrane rafts. iInt. J. Med. MIcrobiol. **291**, 433 (2002). (cited on p. 14)

[30] F. G. van der Goot and T. Harder. Raft membrane domains: from a liquid-ordered membrane phase to a site of pathogen attack. Seminars in Immunology **13**, 89 (2001). (cited on pages 14, 17, and 130)

[31] D. Lingwood and K. Simons. Lipid Rafts As a Membrane-Organizing Principle. Science **327**, 46 (2010). (cited on p. 14)

[32] M. S. Bretscher. Assymetrical Lipid Bilayer Structure for Biological Membranes. Nature **236**, 11 (1972). (cited on p. 14)

[33] P. A. Janmey and P. Kinnunen. Biophysical properties of lipids and dynamic membranes. Trends in Cell Biology (2006). (cited on p. 15)

[34] H. T. Mcmahon and J. L. Gallop. Membrane curvature and mechanisms of dynamic cell membrane remodelling. Nature **438**, 590 (2005). (cited on pages 15 and 130)

[35] M. P. Sheetz and J. Dai. Modulation of membrane dynamics and cell motility by membrane tension. Cell Biology **6**, 85 (1996). (cited on p. 15)

[36] T. Baumgart, S. T. Hess, and W. W. Webb. Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. Nature **425**, 821 (2003). (cited on pages 15 and 97)

[37] R. Lipowsky. Domain-induced budding of fluid membranes. Biophysical Journal **64**, 1133 (1993). (cited on pages 15, 30, and 97)

[38] A. Benmerah and C. Lamaze. Clathrin-Coated Pits: Vive La Différence? Traffic 970–982 (2007). (cited on p. 16)

[39] G. J. Doherty and H. T. Mcmahon. Mechanisms of Endocytosis. Annual Review of Biochemistry **78**, 857 (2009). (cited on pages 16, 17, 97, and 130)

[40] J. Z. Rappoport. Focusing on clathrin-mediated endocytosis. Biochem. J. **412**, 415 (2008). (cited on p. 17)

[41] B. J. Nichols and J. Lippincott-schwartz. Endocytosis without clathrin coats. Trends in Cell Biology **11**, 406 (2001). (cited on p. 17)

[42] A. A. Waheed and E. O. Freed. Lipids and Membrane Microdomains in HIV-1 Replication. Virus. Res. **143**, 162 (2009). (cited on pages 17, 29, 97, and 130)

[43] M. Horn. Chlamydiae as Symbionts in Eukaryotes. Annual Review of Microbiology **62**, 113 (2008). (cited on p. 18)

[44] P. G. Cantalupo, B. Calgua, G. Zhao, R. Sewage, H. Diverse, V. Populations, S. This, R. S. S. Feeds, and A. S. M. Journal. Raw Sewage Harbors Diverse Viral Populations. mBio **2**, 1 (2011). (cited on p. 18)

[45] R. A. Edwards and F. Rohwer. Viral metagenomics. Nature Reviews Microbiology **3**, 504 (2005). (cited on p. 18)

[46] R. Sanjuán, M. R. Nebot, N. Chirico, L. M, and R. Belshaw. Viral Mutation Rates. Journal of Virology **84**, 9733 (2010). (cited on p. 18)

[47] S. Flint, L. Enquist, V. Racaniello, and A. Skalka. *Principles of virology: molecular virology, pathogenesis, and control of animal viruses* (ASM Press, Washington, DC, 2004), 2nd edition. (cited on pages 19, 20, 22, 24, 29, and 34)

[48] L. M. Iyer, S. Balaji, E. V. Koonin, and L. Aravind. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Research **117**, 156 (2006). (cited on p. 19)

[49] D. Prangishvili, R. A. Garrett, and E. V. Koonin. Evolutionary genomics of archaeal viruses: Unique viral genomes in the third domain of life. Virus Research **117**, 52 (2006). (cited on p. 19)

[50] E. M. Tsagris, M. Gozmanova, and K. Kalantidis. Viroids. Cellular Microbiology **10**, 2168 (2008). (cited on p. 19)

[51] S. B. Prusiner. Prions. PNAS **95**, 13363 (1998). (cited on p. 19)

[52] F. H. C. Crick and J. D. Watson. Structure of Small Viruses. Nature **177**, 473 (1956). (cited on p. 20)

[53] K. Namba and G. Stubbs. Structure of Tobacco Mosaic Virus at 3.6 A Resolution: Implications for Assembly. Science **231**, 1401 (1986). (cited on p. 21)

[54] H. Gelderblom. *Structure and Classification of Viruses* (Galveston, TX, 1996), 4th edition. (cited on p. 21)

[55] D. Caspar and A. Klug. Physical Principles in the Construction of Regular Viruses. Cold Spring Harbor Symposium of Quantitative Biology **27**, 1 (1962). (cited on p. 21)

[56] N. Dimmock, A. Easton, and K. Leppard. *Introduction to Modern Virology* (Blackwell Science, 2001), 5th edition. (cited on pages 22, 26, and 27)

[57] R. V. Mannige and C. L. Brooks. Periodic Table of Virus Capsids: Implications for Natural Selection and Design. PLOS one **5**, 1 (2010). (cited on p. 23)

[58] T. C. Mettenleiter. Budding events in herpesvirus morphogenesis. Virus Research **106**, 167 (2004). (cited on p. 27)

[59] G. L. Smith, A. Vanderplasschen, and M. Law. The formation and function of extracellular enveloped vaccinia virus. J. Gen. Virol. **83**, 2915 (2002). (cited on p. 28)

[60] H. Jin, G. P. Leser, J. Zhang, and R. A. Lamb. Influenza virus hemagglutinin and neuraminidase cytoplasmic tails control particle shape **16**, 1236 (1997). (cited on pages 28 and 131)

[61] D. G. Demirov and E. O. Freed. Retrovirus budding. Virus Research **106**, 87 (2004). (cited on pages 28 and 94)

[62] B. J. Chen and R. A. Lamb. Mechanisms for enveloped virus budding: Can some viruses do without an ESCRT? Virology **372**, 221 (2008). (cited on p. 28)

[63] H. Garoff, R. Hewson, and D.-J. E. Opstelten. Virus Maturation by Budding. Microbiology and Molecular Biology Reviews **62**, 1171 (1998). (cited on pages 28, 30, and 94)

[64] L.-A. Carlson, J. A. Briggs, B. Glass, J. D. Riches, and et al Martha N. Simon. Three-Dimensional Analysis of Budding Sites and Released Virus Suggests a Revised Model for HIV-1 Morphogenesis ¨. Cell Host & Microbe **4**, 592 (2008). (cited on pages 29, 31, 98, and 131)

[65] J. A. G. Briggs, J. D. Riches, B. Glass, V. Bartonova, and G. Zanetti. Structure and assembly of immature HIV ¨. PNAS **106**, 11090 (2009). (cited on pages 29, 98, and 131)

[66] A. Ono and E. O. Freed. Role of lipid rafts in virus replication. Advances in Virus Research **64**, 311 (2005). (cited on p. 29)

[67] S. Welsch, O. T. Keppler, A. Habermann, I. Allespach, J. Krijnse-Locker, and H.-G. Kräusslich. HIV-1 buds predominantly at the plasma membrane of primary human macrophages. PLoS pathogens **3**, e36 (2007). (cited on p. 30)

[68] S. Welsch, B. Müller, and H.-G. Kräusslich. More than one door - Budding of enveloped viruses through cellular membranes. FEBS letters **581**, 2089 (2007). (cited on pages 30, 77, and 79)

[69] J. Solon, O. Gareil, P. Bassereau, and Y. Gaudin. Membrane deformations induced by the matrix protein of vesicular stomatitis virus in a minimal system. The Journal of general virology **86**, 3357 (2005). (cited on pages 30, 78, and 130)

[70] H. Vennema, G. J. Godeke, J. W. Rossen, W. F. Voorhout, M. C. Horzinek, D. J. Opstelten, and P. J. Rottier. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. The EMBO journal **15**, 2020 (1996). (cited on p. 30)

[71] R. H. Cheng, R. J. Kuhn, N. H. Olson, M. G. Rossmann, H.-K. Choi, T. J. Smith, and T. S. Baker. Nucleocapsid and glycoprotein organization in an enveloped virus. Cell **80**, 621 (1995). (cited on p. 30)

[72] A. V. Shnyrova, J. Ayllon, I. I. Mikhalyov, E. Villar, J. Zimmerberg, and V. A. Frolov. Vesicle formation by self-assembly of membrane-bound matrix proteins into a fluidlike budding domain. The Journal of cell biology **179**, 627 (2007). (cited on p. 30)

[73] S. Barman, L. Adhikary, A. K. Chakrabarti, C. Bernas, Y. Kawaoka, and D. P. Nayak. Role of Transmembrane Domain and Cytoplasmic Tail Amino Acid Sequences of Influenza A Virus Neuraminidase in Raft Association and Virus Budding. Journal of Virology **78**, 5258 (2004). (cited on p. 30)

[74] Y. Zhang, H. Qian, Z. Love, and E. Barklis. Analysis of the Assembly Function of the Human Immunodeficiency Virus Type 1 Gag Protein Nucleocapsid Domain. J. Virol. **72**, 1782 (1998). (cited on pages 31 and 97)

[75] E. Popova, S. Popov, and H. G. Göttlinger. Human immunodeficiency virus type 1 nucleocapsid p1 confers ESCRT pathway dependence. Journal of virology **84**, 6590 (2010). (cited on pages 31, 78, and 97)

[76] W. Stoeckenius. Structure of the Plasma Membrane: An Electron-Microscope Study. Circulation **26**, 1066 (1962). (cited on p. 33)

[77] M. Adrian, J. Dubochet, J. Lepault, and A. W. McDowall. Cryo-electron microscopy of viruses. Nature **308**, 32 (1984). (cited on p. 33)

[78] W. M. Stanley. ISOLATION OF A CRYSTALLINE PROTEIN POSSESSING THE PROPERTIES OF TOBACCO-MOSAIC VIRUS. Science (New York, N.Y.) **81**, 644 (1935). (cited on p. 33)

[79] J. Watson. The structure of tobacco mosaic virus. Biochimica et Biophysica Acta **13**, 10 (1954). (cited on p. 33)

[80] R. E. Franklin and A. Klug. The splitting of layer lines in X-ray fibre diagrams of helical structures: application to tobacco mosaic virus. Acta Crystallographica **8**, 777 (1955). (cited on p. 33)

[81] J. Kirz, C. Jacobsen, and M. Howells. Soft X-ray microscopes and their biological applications. Quarterly reviews of biophysics **28**, 33 (1995). (cited on p. 33)

[82] M. Chalfie, Y. Tu, G. Euskirchen, W. Ward, and D. Prasher. Green fluorescent protein as a marker for gene expression. Science **263**, 802 (1994). (cited on p. 33)

[83] L. D. Frye and M. Edidin. The rapid intermixing of cell surface antigens after formation of mouse- human heterokaryons. Cell Sci. **7**, 319 (1970). (cited on p. 33)

[84] G. J. Schütz, G. Kada, and H. Schindler. Properties of lipid microdomains in a muscle cell membrane visualized by single molecule microscopy **19**, 892 (2000). (cited on p. 34)

[85] B. Müller, J. Daecke, O. T. Fackler, M. T. Dittmar, H. Zentgraf, and H.-g. Kra. Construction and Characterization of a Fluorescently Labeled Infectious Human Immunodeficiency Virus Type 1 Derivative. Journal of Virology **78**, 10803 (2004). (cited on p. 34)

[86] B. R. Glick, J. J. Pasternak, and C. L. Patten. Molecular biotechnology: principles and applications of recombinant DNA. (2010). (cited on p. 34)

[87] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. Physical Review B **47**, 558 (1993). (cited on p. 35)

[88] P. Carloni, U. Rothlisberger, and M. Parrinello. The Role and Perspective of Ab Initio Molecular Dynamics in the Study of Biological Systems. Accounts of Chemical Research **35**, 455 (2002). (cited on p. 35)

[89] S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein. Coarse grain models and the computer simulation of soft materials. J. Phys: Condens. Matter **16**, 481 (2004). (cited on pages 36 and 37)

[90] K. Katsov and M. Schick. Biological and synthetic membranes: What can be learned from a coarse-grained description? Physics Reports **434**, 113 (2006).
(cited on pages 37 and 39)

[91] M. Deserno. Mesoscopic Membrane Physics: Concepts, Simulations, and Selected Applications. Macromolecular Rapid Communications **30**, 752 (2009).
(cited on pages 37, 67, and 68)

[92] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein. A Coarse Grain Model for Phospholipid Simulations. J. Phys. Chem. B **105**, 4464 (2001).
(cited on pages 37 and 38)

[93] I. R. Cooke and M. Deserno. Solvent-free model for self-assembling fluid bilayer membranes: Stabilization of the fluid phase based on broad attractive tail potentials. Journal of Chemical Physics **123**, 224710 (2005).
(cited on pages 38, 67, 68, 69, 70, 79, 80, and 105)

[94] P. J. Hoogerbrugge and J. M. V. A. Koelman. Simulating Microscopic Hydrodynamic Phenomena with Dissipative Particle Dynamics. Europhysics Letters (EPL) **19**, 155 (1992). (cited on p. 38)

[95] P. Español and P. Warren. Statistical Mechanics of Dissipative Particle Dynamics. Europhysics Letters (EPL) **30**, 191 (1995). (cited on p. 38)

[96] R. Koynova and M. Caffrey. Phases and phase transitions of the phosphatidylcholines. Biochimica et Biophysica Acta **1376**, 91 (1998). (cited on p. 39)

[97] J. Stümpel, H. Eibl, and A. Nicksch. X-ray analysis and calorimetry on phosphatidylcholine model membranes. The influence of length and position of acyl chains upon structure and phase behaviour. Biochimica et Biophysica Acta (BBA) - Biomembranes **727**, 246 (1983). (cited on p. 39)

[98] H. I. Petrache, S. W. Dodd, and M. F. Brown. Area per lipid and acyl length distributions in fluid phosphatidylcholines determined by (2)H NMR spectroscopy. Biophysical journal **79**, 3172 (2000). (cited on p. 39)

[99] M. Kranenburg, M. Venturoli, and B. Smit. Molecular simulations of mesoscopic bilayer phases. Physical Review E **67**, 060901 (2003). (cited on pages 39 and 66)

[100] M. Kranenburg, M. Venturoli, and B. Smit. Phase Behavior and Induced Interdigitation in Bilayers Studied with Dissipative Particle Dynamics. The Journal of Physical Chemistry B **107**, 11491 (2003). (cited on pages 39 and 66)

[101] R. Faller and S.-J. Marrink. Simulation of domain formation in DLPC-DSPC mixed bilayers. Langmuir : the ACS journal of surfaces and colloids **20**, 7686 (2004). (cited on p. 39)

[102] J. Hjort Ipsen, G. Karlström, O. Mourtisen, H. Wennerström, and M. Zuckermann. Phase equilibria in the phosphatidylcholine-cholesterol system. Biochimica et Biophysica Acta (BBA) - Biomembranes **905**, 162 (1987). (cited on p. 39)

[103] H. I. Petrache, A. Grossfield, K. R. MacKenzie, D. M. Engelman, and T. B. Woolf. Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: molecular dynamics calculations. Journal of molecular biology **302**, 727 (2000). (cited on p. 39)

[104] M. Venturoli, B. Smit, and M. M. Sperotto. Simulation studies of protein-induced bilayer deformations, and lipid-induced protein tilting, on a mesoscopic model for lipid bilayers with embedded proteins. Biophysical journal **88**, 1778 (2005). (cited on pages 39 and 66)

[105] C. F. Lopez, S. O. Nielsen, P. B. Moore, and M. L. Klein. Understanding nature's design for a nanosyringe. Proceedings of the National Academy of Sciences of the United States of America **101**, 4431 (2004). (cited on p. 39)

[106] C. F. Lopez, S. O. Nielsen, B. Ensing, P. B. Moore, and M. L. Klein. Structure and dynamics of model pore insertion into a membrane. Biophysical journal **88**, 3083 (2005). (cited on p. 39)

[107] H. Noguchi and M. Takasu. Self-assembly of amphiphiles into vesicles: A Brownian dynamics simulation. Physical Review E **64** (2001). (cited on pages 39 and 66)

[108] M. Müller, K. Katsov, and M. Schick. Coarse-grained models and collective phenomena in membranes: Computer simulation of membrane fusion. Journal of Polymer Science Part B: Polymer Physics **41**, 1441 (2003). (cited on p. 39)

[109] M. Venturoli, M. M. Sperotto, M. Kranenburg, and B. Smit. Mesoscopic models of biological membranes. Physics Reports **437**, 1 (2006). (cited on pages 39, 64, and 67)

[110] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1990). (cited on pages 40, 41, 43, and 55)

[111] D. Frenkel and B. Smit. *Understanding Molecular Simulations. From Algorithms to Applications* (Academic Press, 2002). (cited on pages 40, 42, 47, 57, and 81)

[112] S. H. Strogatz. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry And Engineering (Studies in Nonlinearity)* (Westview Press, 1994). (cited on p. 41)

[113] H. Goldstein, C. P. Poole, and J. L. Safko. *Classical Mechanics* (Addison Wesley, 2001). (cited on p. 44)

[114] J. Barojas, D. Levesque, and B. Quentrec. Simulation of Diatomic Homonuclear Liquids. Physical Review A **7**, 1092 (1973). (cited on p. 44)

[115] T. D. Nguyen, C. L. Phillips, J. a. Anderson, and S. C. Glotzer. Rigid body constraints realized in massively-parallel molecular dynamics on graphics processing units. Computer Physics Communications **182**, 2307 (2011). (cited on pages 46 and 113)

[116] W. van Gunsteren and H. Berendsen. Algorithms for brownian dynamics. Molecular Physics **45**, 637 (1982). (cited on p. 49)

[117] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. The Journal of Chemical Physics **72**, 2384 (1980). (cited on p. 49)

[118] M. Parrinello and A. Rahman. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. Physical Review Letters **45**, 1196 (1980). (cited on p. 49)

[119] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant pressure molecular dynamics algorithms. J. Chem. Phys. **101**, 4177 (1994). (cited on pages 49 and 112)

[120] A. Kolb and B. Dünweg. Optimized constant pressure stochastic dynamics. Chemical Physics **111**, 4453 (1999). (cited on pages 50 and 81)

[121] G. Torrie and J. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. Journal of Computational Physics **23**, 187 (1977). (cited on p. 53)

[122] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. Journal of Computational Chemistry **13**, 1011 (1992). (cited on p. 54)

[123] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: fast, flexible, and free. Journal of computational chemistry **26**, 1701 (2005). (cited on p. 58)

[124] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. Scalable molecular dynamics with NAMD. Journal of computational chemistry **26**, 1781 (2005). (cited on p. 58)

[125] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry **4**, 187 (1983). (cited on p. 58)

[126] J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco, and K. Schulten. Accelerating Molecular Modeling Applications with Graphics Processors. J. Comput. Chem. **28**, 2618 (2007). (cited on p. 59)

[127] CUDA. (cited on p. 59)

[128] J. A. Anderson, C. D. Lorenz, and A. Travesset. General purpose molecular dynamics simulations fully implemented on graphics processing units. Journal of Computational Physics **227**, 5342 (2008). (cited on p. 59)

[129] HOOMD-blue. (cited on pages 59 and 112)

[130] W. Helfrich. Elastic properties of lipid bilayers: theory and possible experiments. Zeitschrift für Naturforschung. Teil C: Biochemie, Biophysik, Biologie, Virologie **28**, 693 (1973). (cited on pages 59 and 83)

[131] M. P. D. Carmo. *Differential Geometry of Curves and Surfaces* (Prentice Hall, 1976). (cited on pages 60 and 83)

[132] V. A. Harmandaris and M. Deserno. A novel method for measuring the bending rigidity of model lipid membranes by simulating tethers. The Journa **125**, 1 (2006). (cited on p. 61)

[133] a. Imparato, J. C. Shillcock, and R. Lipowsky. Shape fluctuations and elastic properties of two-component bilayer membranes. Europhysics Letters (EPL) **69**, 650 (2005). (cited on p. 61)

[134] W. K. den Otter and W. J. Briels. The bending rigidity of an amphiphilic bilayer from equilibrium and nonequilibrium molecular dynamics. The Journal of Chemical Physics **118**, 4712 (2003). (cited on p. 61)

[135] A. Okabe, B. Boots, K. Sugihara, D. S. N. Chiu, and S. N. Chiu. *Spatial tessellations: Concepts and applications of Voronoi diagrams* (Wiley, 2000), 2nd edition. (cited on p. 61)

[136] R. Goetz and R. Lipowsky. Computer simulations of bilayer membranes: Self-assembly and interfacial tension ¨. Journal of Chemical Physics **108**, 7397 (1998). (cited on pages 64 and 110)

[137] W. K. den Otter and W. J. Briels. The bending rigidity of an amphiphilic bilayer from equilibrium and nonequilibrium molecular dynamics. The Journal of Chemical Physics **118**, 4712 (2003). (cited on p. 64)

[138] M. J. Stevens. Coarse-grained simulations of lipid bilayers. The Journal of chemical physics **121**, 11942 (2004). (cited on pages 65 and 66)

[139] M. J. Stevens. Complementary matching in domain formation within lipid bilayers. Journal of the American Chemical Society **127**, 15330 (2005). (cited on p. 65)

[140] S. J. Marrink, A. H. D. Vries, and A. E. Mark. Coarse Grained Model for Semiquantitative Lipid Simulations. Journal of Physical Chemistry B **108**, 750 (2004). (cited on p. 65)

[141] S. J. Marrink, J. Risselada, and A. E. Mark. Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model. Chemistry and physics of lipids **135**, 223 (2005). (cited on p. 65)

[142] S. J. Marrink and A. E. Mark. Molecular Dynamics Simulation of the Formation, Structure, and Dynamics of Small Phospholipid Vesicles. Analysis 15233–15242 (2003). (cited on pages 65 and 66)

[143] S.-J. Marrink and A. E. Mark. Molecular view of hexagonal phase formation in phospholipid membranes. Biophysical journal **87**, 3894 (2004). (cited on p. 65)

[144] M. Kranenburg and B. Smit. Simulating the effect of alcohol on the structure of a membrane. FEBS letters **568**, 15 (2004). (cited on p. 66)

[145] S. Izvekov and G. A. Voth. Multiscale Coarse-Graining of Mixed Phospholipid/Cholesterol Bilayers. Journal of Chemical Theory and Computation **2**, 637 (2006). (cited on p. 66)

[146] S. O. Nielsen, C. F. Lopez, I. Ivanov, P. B. Moore, J. C. Shelley, and M. L. Klein. Transmembrane peptide-induced lipid sorting and mechanism of Lalpha-to-inverted phase transition using coarse-grain molecular dynamics. Biophysical journal **87**, 2107 (2004). (cited on p. 66)

[147] P. M. Kasson, N. W. Kelley, N. Singhal, M. Vrljic, A. T. Brunger, and V. S. Pande. Ensemble molecular dynamics yields submillisecond kinetics and intermediates of membrane fusion. Proceedings of the National Academy of Sciences of the United States of America **103**, 11916 (2006). (cited on p. 66)

[148] T. Sintes and A. Baumgärtner. Protein attraction in membranes induced by lipid fluctuations. Biophysical journal **73**, 2251 (1997). (cited on p. 66)

[149] M. Laradji. A Monte Carlo study of fluctuating polymer-grafted membranes. The Journal of chemical physics **121**, 1591 (2004). (cited on p. 66)

[150] J. Drouffe, A. Maggs, and S. Leibler. Computer simulations of self-assembled membranes. Science **254**, 1353 (1991). (cited on p. 66)

[151] G. Brannigan, L. C.-L. Lin, and F. L. H. Brown. Implicit solvent simulation models for biomembranes. European biophysics journal : EBJ **35**, 104 (2006). (cited on pages 66 and 67)

[152] H. Noguchi and M. Takasu. Adhesion of nanoparticles to vesicles: a Brownian dynamics simulation. Biophysical journal **83**, 299 (2002). (cited on p. 66)

[153] O. Farago and I. Introduction. "Water-free" computer model for fluid bilayer membranes. Chemical Physics **119**, 596 (2003). (cited on p. 67)

[154] O. Farago and C. D. Santangelo. Pore formation in fluctuating membranes. The Journal of chemical physics **122**, 44901 (2005). (cited on p. 67)

[155] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. The Journal of Chemical Physics **54**, 5237 (1971). (cited on pages 68 and 80)

[156] G. Grest and K. Kremer. Molecular dynamics simulation for polymers in the presence of a heat bath. Physical Review A **33**, 3628 (1986). (cited on p. 68)

[157] I. R. Cooke and M. Deserno. Coupling between lipid shape and membrane curvature. Biophysical journal **91**, 487 (2006). (cited on p. 69)

[158] B. J. Reynwar, G. Illya, V. A. Harmandaris, M. M. Müller, K. Kremer, and M. Deserno. Aggregation and vesiculation of membrane proteins by curvature-mediated interactions. Nature **447**, 461 (2007). (cited on pages 69 and 115)

[159] M. I. Hoopes, M. Deserno, M. L. Longo, and R. Faller. Coarse-grained modeling of interactions of lipid bilayers with supports. The Journal of chemical physics **129**, 175102 (2008). (cited on p. 69)

[160] H. Fraenkel-Conrat and R. C. Williams. Reconstitution of active tocabbo mosaic virus from its inactive protein and nucleic acid components. Proceedings of the National Academy of Sciences of the United States of America **41**, 690 (1955). (cited on p. 70)

[161] Hagan. virus review . (cited on pages 70, 71, 72, 98, and 100)

[162] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. Structure (London, England : 1993) **14**, 437 (2006). (cited on p. 71)

[163] G. S. Ayton and G. A. Voth. Multiscale computer simulation of the immature HIV-1 virion. Biophysical journal **99**, 2757 (2010). (cited on p. 71)

[164] M. F. Hagan and D. Chandler. Dynamic pathways for viral capsid assembly. Biophysical journal **91**, 42 (2006). (cited on pages 71 and 72)

[165] R. Schwartz, P. W. Shor, P. E. Prevelige, and B. Berger. Local rules simulation of the kinetics of virus capsid self-assembly. Biophysical journal **75**, 2626 (1998). (cited on p. 71)

[166] A. W. Wilber, J. P. K. Doye, A. A. Louis, E. G. Noya, M. A. Miller, and P. Wong. Reversible self-assembly of patchy particles into monodisperse icosahedral clusters. The Journal of chemical physics **127**, 085106 (2007). (cited on p. 71)

[167] D. Rapaport. Self-assembly of polyhedral shells: A molecular dynamics study. Physical Review E **70** (2004). (cited on pages 71 and 72)

[168] A. Levandovsky and R. Zandi. Nonequilibirum assembly, retroviruses, and conical structures. Physical review letters **102**, 198102 (2009). (cited on pages 71 and 72)

[169] D. C. Rapaport. Modeling capsid self-assembly: design and analysis. Physical biology **7**, 045001 (2010). (cited on p. 71)

[170] H. D. Nguyen, V. S. Reddy, and C. L. Brooks. Invariant polymorphism in virus capsid assembly. Journal of the American Chemical Society **131**, 2606 (2009). (cited on p. 71)

[171] M. F. Hagan, O. M. Elrad, and R. L. Jack. Mechanisms of kinetic trapping in self-assembly and phase transformation. The Journal of chemical physics **135**, 104115 (2011). (cited on p. 71)

[172] O. M. Elrad and M. F. Hagan. Encapsulation of a polymer by an icosahedral virus. Physical biology **7**, 045003 (2010). (cited on p. 71)

[173] H. D. Nguyen, V. S. Reddy, and C. L. Brooks. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. Nano letters **7**, 338 (2007). (cited on p. 71)

[174] D. Rapaport, J. Johnson, and J. Skolnick. Supramolecular self-assembly: molecular dynamics modeling of polyhedral shell formation. Computer Physics Communications **121-122**, 231 (1999). (cited on p. 71)

[175] J. P. Mahalik and M. Muthukumar. Langevin dynamics simulation of polymer-assisted virus-like assembly. The Journal of chemical physics **136**, 135101 (2012). (cited on p. 71)

[176] S. Hicks and C. Henley. Irreversible growth model for virus capsid assembly. Physical Review E **74**, 31912 (2006). (cited on p. 71)

[177] R. Bruinsma, W. Gelbart, D. Reguera, J. Rudnick, and R. Zandi. Viral Self-Assembly as a Thermodynamic Process. Physical Review Letters **90** (2003). (cited on p. 72)

[178] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick. Origin of icosahedral symmetry in viruses. Proceedings of the National Academy of Sciences of the United States of America **101**, 15556 (2004). (cited on p. 72)

[179] A. Luque, R. Zandi, and D. Reguera. Optimal architectures of elongated viruses. Proceedings of the National Academy of Sciences of the United States of America **107**, 5323 (2010). (cited on p. 72)

[180] D. J. Wales. The energy landscape as a unifying theme in molecular science. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences **363**, 357 (2005). (cited on pages 73, 75, 99, and 130)

[181] S. N. Fejer, T. R. James, J. Hernández-Rojas, and D. J. Wales. Energy landscapes for shells assembled from pentagonal and hexagonal pyramids. Physical chemistry chemical physics : PCCP **11**, 2098 (2009). (cited on pages 73, 74, 75, and 105)

[182] I. G. Johnston, A. A. Louis, and J. P. K. Doye. Modelling the self-assembly of virus capsids. Journal of physics. Condensed matter : an Institute of Physics journal **22**, 104101 (2010). (cited on pages 75 and 105)

[183] A. E. Nel, L. Mädler, D. Velegol, T. Xia, E. M. V. Hoek, P. Somasundaran, F. Klaessig, V. Castranova, and M. Thompson. Understanding biophysicochemical interactions at the nano-bio interface. Nature materials **8**, 543 (2009). (cited on p. 77)

[184] S. Mitragotri and J. Lahann. Physical approaches to biomaterial design. Nature materials **8**, 15 (2009). (cited on p. 77)

[185] C. A. Poland, R. Duffin, I. Kinloch, A. Maynard, W. A. H. Wallace, A. Seaton, V. Stone, S. Brown, W. Macnee, and K. Donaldson. Carbon nanotubes introduced into the abdominal cavity of mice show asbestos-like pathogenicity in a pilot study. Nature nanotechnology **3**, 423 (2008). (cited on p. 77)

[186] J. Liu, Y. Sun, D. G. Drubin, and G. F. Oster. The mechanochemistry of endocytosis. PLoS biology **7**, e1000204 (2009). (cited on pages 77 and 95)

[187] M. Gladnikoff, E. Shimoni, N. S. Gov, and I. Rousso. Retroviral assembly and budding occur through an actin-driven mechanism. Biophysical journal **97**, 2419 (2009). (cited on pages 77 and 94)

[188] V. Baumgärtel, S. Ivanchenko, A. Dupont, M. Sergeev, P. W. Wiseman, H.-G. Kräusslich, C. Bräuchle, B. Müller, and D. C. Lamb. Live-cell visualization of dynamics of HIV budding site interactions with an ESCRT component. Nature cell biology **13**, 469 (2011). (cited on pages 78, 94, and 97)

[189] A. Helenius, E. Fries, and J. Kartenbeck. Reconstitution of Semliki forest virus membrane. The Journal of cell biology **75**, 866 (1977). (cited on p. 78)

[190] C. H. von Bonsdorff and S. C. Harrison. Hexagonal glycoprotein arrays from Sindbis virus membranes. J. Virol. **28**, 578 (1978). (cited on p. 78)

[191] O. Le Bihan, P. Bonnafous, L. Marak, T. Bickel, S. Trépout, S. Mornet, F. De Haas, H. Talbot, J.-C. Taveau, and O. Lambert. Cryo-electron tomography of nanoparticle transmigration into liposome. Journal of structural biology **168**, 419 (2009). (cited on p. 78)

[192] J. H. Hurley, E. Boura, L.-A. Carlson, and B. Różycki. Membrane budding. Cell **143**, 875 (2010). (cited on pages 78, 97, and 130)

[193] J. S. Rossman, X. Jing, G. P. Leser, and R. A. Lamb. Influenza virus M2 protein mediates ESCRT-independent membrane scission. Cell **142**, 902 (2010). (cited on pages 78 and 97)

[194] M. Deserno and W. M. Gelbart. Adhesion and Wrapping in Colloid-Vesicle Complexes. The Journal of Physical Chemistry B **106**, 5543 (2002). (cited on pages 78, 82, 83, 84, and 93)

[195] H. Gao, W. Shi, and L. B. Freund. Mechanics of receptor-mediated endocytosis. Proceedings of the National Academy of Sciences of the United States of America **102**, 9469 (2005). (cited on p. 78)

[196] R. Zhang and T. Nguyen. Model of human immunodeficiency virus budding and self-assembly: Role of the cell membrane. Physical Review E **78** (2008). (cited on pages 78, 93, and 98)

[197] A. Chaudhuri, G. Battaglia, and R. Golestanian. The effect of interactions on the cellular uptake of nanoparticles. Physical biology **8**, 046002 (2011). (cited on pages 78 and 94)

[198] M. Fošnariš, A. Iglicš, D. M. Kroll, and S. May. Monte Carlo simulations of complex formation between a mixed fluid vesicle and a charged colloid. The Journal of Chemical Physics **131**, 105103 (2009). (cited on p. 78)

[199] Y. Li and N. Gu. Thermodynamics of charged nanoparticle adsorption on charge-neutral membranes: a simulation study. The journal of physical chemistry. B **114**, 2749 (2010). (cited on p. 78)

[200] V. V. Ginzburg and S. Balijepalli. Modeling the thermodynamics of the interaction of nanoparticles with cell membranes. Nano letters **7**, 3716 (2007). (cited on p. 78)

[201] K. A. Smith, D. Jasnow, and A. C. Balazs. Designing synthetic vesicles that engulf nanoscopic particles. The Journal of chemical physics **127**, 084703 (2007). (cited on pages 78, 88, and 109)

[202] T. Yue and X. Zhang. Molecular understanding of receptor-mediated membrane responses to ligand-coated nanoparticles. Soft Matter **7**, 9104 (2011). (cited on pages 78, 89, and 93)

[203] K. Yang and Y.-q. Ma. Wrapping and Internalization of Nanoparticles by Lipid Bilayers: a Computer Simulation Study. Australian Journal of Chemistry **64**, 894 (2011). (cited on p. 78)

[204] R. Vácha, F. J. Martinez-Veracoechea, and D. Frenkel. Receptor-mediated endocytosis of nanoparticles of various shapes. Nano letters **11**, 5391 (2011). (cited on p. 78)

[205] H. Garoff, M. Sjöberg, and R. H. Cheng. Budding of alphaviruses. Virus research **106**, 103 (2004). (cited on pages 79 and 94)

[206] A. Grossfield. WHAM:the weighted histogram analysis method. (cited on p. 82)

[207] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. Journal of molecular graphics **14**, 33 (1996). (cited on p. 87)

[208] Y. Li, X. Chen, and N. Gu. Computational investigation of interaction between nanoparticles and membranes: hydrophobic/hydrophilic effect. The journal of physical chemistry. B **112**, 16647 (2008). (cited on p. 89)

[209] C. Dellago, P. Bolhuis, and P. Geissler. *Advances in Chemical Physics*, volume 123 (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2002). (cited on p. 91)

[210] W. Jiang, B. Y. S. Kim, J. T. Rutka, and W. C. W. Chan. Nanoparticle-mediated cellular response is size-dependent. Nature nanotechnology **3**, 145 (2008). (cited on p. 94)

[211] C. Seeger and W. S. Mason. Hepatitis B virus biology. Microbiology and molecular biology reviews : MMBR **64**, 51 (2000). (cited on p. 94)

[212] A. R. Honerkamp-Smith, P. Cicuta, M. D. Collins, S. L. Veatch, M. den Nijs, M. Schick, and S. L. Keller. Line tensions, correlation lengths, and critical exponents in lipid membranes near critical points. Biophysical journal **95**, 236 (2008). (cited on p. 94)

[213] A. Ono and E. O. Freed. Plasma membrane rafts play a critical role in HIV-1 assembly and release. Proceedings of the National Academy of Sciences of the United States of America **98**, 13925 (2001). (cited on pages 94 and 97)

[214] M. P. Taylor, O. O. Koyuncu, and L. W. Enquist. Subversion of the actin cytoskeleton during viral infection. Nature reviews. Microbiology **9**, 427 (2011). (cited on pages 94, 97, and 130)

[215] R. B. Chan, L. Tanner, and M. R. Wenk. Implications for lipids during replication of enveloped viruses. Chemistry and physics of lipids **163**, 449 (2010). (cited on p. 94)

[216] H. R. Jayakar, E. Jeetendra, and M. A. Whitt. Rhabdovirus assembly and budding. Virus Research **106**, 117 (2004). (cited on p. 97)

[217] D. P. Nayak, E. K.-W. Hui, and S. Barman. Assembly and budding of influenza virus. Virus research **106**, 147 (2004). (cited on p. 97)

[218] E. O. Freed and A. J. Mouland. The cell biology of HIV-1 and other retroviruses. Retrovirology **3**, 1 (2006). (cited on p. 97)

[219] N. Jouvenet, M. Zhadina, P. D. Bieniasz, and S. M. Simon. Dynamics of ESCRT protein recruitment during retroviral assembly. Nature cell biology **13**, 394 (2011). (cited on p. 97)

[220] S. Ivanchenko, W. J. Godinez, M. Lampe, H.-G. Kräusslich, R. Elis, C. Bräuchle, B. Müller, D. C. Lamb, and C. Bra. Dynamics of HIV-1 Assembly and Release. PLoS Pathogens **5** (2009). (cited on pages 97 and 132)

[221] F. Jülicher and R. Lipowsky. Domain-induced budding of vesicles. Physical Review Letters **70**, 2964 (1993). (cited on p. 97)

[222] M. Laradji and P. Sunil Kumar. Dynamics of Domain Growth in Self-Assembled Fluid Vesicles. Physical Review Letters **93** (2004). (cited on p. 97)

[223] N. Jouvenet, P. D. Bieniasz, and S. M. Simon. Imaging the biogenesis of individual HIV-1 virions in live cells. Nature **454**, 236 (2008). (cited on p. 97)

[224] R. Matthews and C. Likos. Influence of Fluctuating Membranes on Self-Assembly of Patchy Colloids. Physical Review Letters **109**, 178302 (2012). (cited on pages 98 and 131)

[225] E. Hamard-Peron and D. Muriaux. Retroviral matrix and lipids, the intimate interaction. Retrovirology **8**, 15 (2011). (cited on p. 104)

[226] S. D. Bond, B. J. Leimkuhler, and B. B. Laird. The Nosé–Poincaré Method for Constant Temperature Molecular Dynamics. Journal of Computational Physics **151**, 114 (1999). (cited on p. 112)

[227] C. T. Jones, L. Ma, J. W. Burgner, T. D. Groesch, C. B. Post, and R. J. Kuhn. Flavivirus Capsid Is a Dimeric Alpha-Helical Protein. Journal of Virology **77**, 7143 (2003). (cited on p. 132)